Technical Report 316

Hypothesis Formation and Evaluation in Medical Diagnosis

Ann D. Rubin

MIT Artificial Intelligence Laboratory

This blank page was inserted to preserve pagination.

Revised version of a dissertation submitted to the Department of Electrical Engineering to the Faculty of Stanford University in partial fulfillment of the requirements for the degree of Master of Science.

ABSTRACT

This thesis describes some aspects of a computer system for doing medical diagnosis in the specialized field of kidney disease. Because such a system faces the spectre of combinatorial explosion, this discussion concentrates on heuristics which control the number of concurrent hypotheses and efficient "compiled" representations of medical knowledge.

In particular, the differential diagnosis of hematuria (blood in the urine) is discussed in detail. A protocol of a simulated doctor/patient interaction is presented and analyzed to determine the crucial structures and processes involved in the diagnostic procedure. The data structure proposed for representing medical information revolves around elementary hypotheses which are activated when certain key findings are discovered. A four-step process which consists of disposing of findings, activating hypotheses, evaluating hypotheses locally and combining hypotheses globally is examined for its heuristic implications.

The thesis attempts to fit the problem of medical diagnosis into the framework of other Artificial Intelligence problems and paradigms and in particular explores the notions of pure search vs. heuristic methods, linearity and interaction, local vs. global knowledge and the structure of hypotheses within the world of kidney disease.

Thesis Supervisor: Gerald J. Sussman
Assistant Professor of Electrical Engineering

Acknowledgements

I would like to thank the following people, without whose help this work would never have attained its current state.

Gerry Sussman, for being a most patient, helpful and diligent thesis advisor, especially at difficult times.

Tony Gorry, for providing me the opportunity to do this research, for being a brave project director and for bringing together doctors and computer scientists in what promises to be a successful coalition.

Dr. J. P. Kassirer, for being an invaluable source of medical knowledge and for trying valiantly to understand my theories.

Dr. Steve Pauker, for doing much of the work this thesis draws from, for providing keen intellectual stimulation and for, hopefully, carrying this work much further.

Ira Goldstein, for being willing to read unfinished and unedited drafts and to listen to complaints.

Terry Winograd, for seeing eye to eye with me on so many occasions.

The MIT AI Lab, for providing fertile intellectual ground.

Ellis Cohen, Brian Smith, Suzin Jabari, Allison Platt and the inhabitants of 20 Prescott St. for putting up with me.

Chapter 1 - Introduction

Doing research which involves writing a program, inventing a formalism or designing a system to accomplish some task is an activity which can be viewed in two very different lights. Its most immediate goal is to produce a working program or simulation, which may be used in speech understanding, scene analysis, game-playing or medical diagnosis. This more immediate point of view is the one more often discussed in papers, which report on a finished or soon-to-be-finished product. From an Artificial Intelligence point of view, however, it is more important to consider the problem-solving process as an exploration of alternative approaches to representation and control structure, as the instantiation or discovery of more general concepts and theories, whose details are of lesser importance. This perspective has been particularly emphasized in AI, a field whose goal is to investigate general problem-solving strategies and wide-ranging insights into possible patterns of human thought.

This thesis studies the problem of medical diagnosis basically from the second point of view, although it recognizes the necessity of paying attention to some of the details in any complex problem domain. It attempts to fit the problem of medical diagnosis into the framework of other AI problems and paradigms and in particular explores the notions of pure search vs. heuristic methods, linearity and interaction, plausibility and the structure of hypotheses within the

not-so-mini-world of kidney disease.

1.1 Why Medicine?

The practical importance of studying and developing computer aids for medical diagnosis is obvious. Doctors train for years to become expert diagnosticians; they carry heavy responsibility for the accuracy of their diagnoses and the effectiveness of their treatments. Yet with all their training, they often make mistakes because of the vast body of ever-increasing medical knowledge they must remember and access. In a computer, the problem of pure memory disappears, while effort focuses instead on methods of representation of knowledge, selection of relevant knowledge and proper use of the selected facts.

Several diagnosis programs have already been written for small areas of medicine such as bone tumors <Gorry 67> and acute renal failure <Gorry 73>; a group at Rutgers is currently analyzing the time course of glaucoma and using their model to place a patient at a point along the temporal progression of the disease and thus determine the prescribed treatment. <Amarel 73> Programs have been written as well to investigate treatment choices <Schwartz 74> and as clinical aids in prescribing and adjusting antibiotic therapies. <Shortliffe 74> <Silverman 74> is currently working on making a program to calculate digitalis doses more sensitive to the individual patient and capable of using his or her reaction to the initial dose to revise its

suggestions. These researchers envision the ultimate use of their programs to be in aiding doctors and augmenting their knowledge, as opposed to replacing them. In the imagined future, GP's will be able to consult a computer for expert advice in areas in which a general practitioner is necessarily less knowledgeable than a specialist. <Schwartz 70> contains a fuller discussion of such future scenarios.

More recent medical diagnosis programs attempt to deal with wider varieties and larger numbers of diseases, to offer coherent explanations of diagnoses, and are based on models of the time course of diseases. In addition, there has been growing interest in the psychological processes of hypothesis-generation and decision-making in medical practice. Medical educators envision this leading to better instruction for students in diagnostic skills, data organization, and test selection.

Another group, the cognitive psychologists and AI researchers, are interested in the structure of medical knowledge and the processes by which it is manipulated as examples of general knowledge structures and problem-solving processes. Medicine has many characteristics which make it well-suited for such theoretical exploration:

COMPLEXITY AND RICHNESS

1. There is no question that medical diagnosis is a complex and rich domain. Certainly, the data itself seems to be complicated (or at least massive) and even a cursory glance at the kind and amount of processing which must occur is enough to justify studying it

further. That there is some kind of rich structure present at least in many doctors' minds, if not in the data itself, is evident if we assume that diagnostic and question-asking strategies proceed from the same data structure; no overly-simple structure will account for the complexities of that process. Of course, AI may flounder in domains with too much complexity. Several of the points below suggest that medical diagnosis occupies a favorable spot along the dimension of complexity.

EVALUABILITY

2. The final goal of a medical diagnosis system is clear, at least on one level; we want a program which will produce the "correct" diagnosis (i.e. the same one as an "expert" would arrive at) at the end of some reasonable amount of processing. This is in contrast to the problem of defining "understanding" in a (language) understanding system. Many attempts have been made to come up with a taxonomy of the indicators of understanding <Newell 73> <Card 74>, but the problem is not a small one and no one would claim it has been satisfactorily solved. On the other hand, we notice that automatic programming problems do have a more clearly-defined goal: the production of a program which performs according to some externally-stated standards; many problems still exist, though, in defining languages in which to state those standards. Of course, in both medicine and debugging, it is the process of arriving at the solution in which we are ultimately interested and the standards for judging these processes are much less

well-specified or understood (but see below, #3). Still, we have at least a first-order criterion by which to judge diagnostic programs.

ACCESS TO INTERMEDIATE RESULTS AND "PROTOCOLS"

3. As mentioned above, process is of primary interest in looking at problem-solving programs; one problem which many theories of problem-solving have had is that there was a lack of natural data giving insight into that process. Most AI programs have tackled one of three major areas: the synthesis of visual scenes from primitive data, the understanding of simple English dialogue and the solution or study of mathematical and other "puzzles," including games like chess and checkers. The "success" of most problem-solving theories developed in these domains had to be judged by a comparison of its results with the "correct" results - and independently by some general criteria about plausible processes. In visual recognition or language understanding, for example, there are no intermediate points in the process about which people naturally verbalize or to which we have any other access. In the medical diagnosis process, on the other hand, practitioners often verbalize spontaneously; getting informal protocols requires only sitting in on clinical sessions or listening to discussions on rounds. More formal and complete protocols are also easily obtainable, since public diagnostic sessions and CPC's (see section 1.2) are common occurrences in hospitals. In this respect, studying medical diagnosis contrasts with taking protocols of subjects solving cryptarithmetic problems, <Newell and Simon 72> which uses an

artificial task in an artificial situation, as well as with language understanding or visual scene analysis, to whose decision processes we have no natural access. (Of course, we must be cautious in our interpretations of protocols as exactly reflecting the reasoning process the physician is using. Section 2.1 considers the significance of protocols in this research and their relationship to the underlying thought processes.)

TERMINOLOGICAL CONCEPTS AND PRIMITIVES

4. Medicine contrasts with vision, although both have been treated as recognition problems (see section 1.2), in terms of the vocabulary available for each subject area. Much of the work which has gone into current vision systems has been devoted to coming up with a limited yet sufficient vocabulary to describe structures as simple as vertices and angles and as complex as textures, curves and complex shapes. <Fahlman 73a> <Hollerbach 74> Medicine, on the other hand, comes completely equipped with a large technical (and sometimes baroque) vocabulary, whose stated aim is, in fact, to allow exact and accurate communication among doctors. Thus, a lot of effort has already been devoted to making the necessary distinctions among symptoms and disease states. We have, unfortunately, found that medical vocabulary is sometimes more confused than one would hope - definitions may be unclear and diseases may overlap. The basic structure, however, has already been laid down.

POSSIBLE MINI-WORLDS

5. Medical diagnosis is so large and varied a field that it allows the construction of many different mini-worlds, the exploration of each aiming toward the clarification of different issues. Thus a problem we often face in AI, that of finding an area small enough to study completely, yet large enough to provide real challenge, seems to be well addressed by the choice of medical diagnosis. The subject matter in medicine can be cut along many different dimensions; most often it has been limited by the selection of a small class of diseases, tests and symptoms, as well as by focussing attention on the final diagnosis to the exclusion of process. In addition, complicating issues not specific to medicine such as the representation of time were often excluded or dealt with using special ad hoc mechanisms. For example, the Rutgers group has limited their investigation to one disease - glaucoma - and is concentrating instead on determining the stage of the disease which a patient manifests; thus the time course of the disease is specifically and exclusively considered. <Amarel 73> Gorry, on the other hand, chose a larger class of possible diagnoses and handled the time of occurrence of symptoms as one example of a general concept of interaction between symptoms. <Gorry 67> This is not to suggest that the hard problem of modularization has been solved in the case of medical diagnosis - but merely to inject some hope; the sub-domains are there, if we can only

find and isolate them.

1.2 Description of the Problem

The particular aspect of medicine with which this thesis will deal is the process of diagnosis within a limited set of diseases: those whose presenting symptom is hematuria, or blood in the urine. We can conceptualize the problem as one of a class of recognition problems <Fahlman 73b> in which features of the situation (called the sample by Fahlman) act as clues to its complete description - to its recognition as an already-known entity. In particular, a medical system is presented with a group of symptoms, signs, facts, test results etc. and its job is to come up with a diagnosis, an identification of a disease or several diseases whose manifestations most closely match the condition of the patient. Choosing a treatment on the basis of the diagnosis will not be included in the analysis here.

Because of our interest in process, the model of diagnosis which will be used here is one of the serial acquisition <Gorry 67> of facts about the patient. Thus, we require a diagnosis system to have hypotheses at each moment and expect that these hypotheses will change after the addition of each new piece of information. As a first approximation, a hypothesis can be thought of as a proposed disease, but several examples later will make it clear that the structure of a

hypothesis is more complicated, often including several related or independent diseases or mechanisms, some of which are connected by relationships like CAUSED-BY or COMPLICATED-BY.

A distinction is often made between two forms of data acquistion in diagnosis: active and passive. <Gorry 74> An active approach includes a physician's asking a question in order to solicit each new piece of information from a patient; clearly his or her questions will rely heavily on the previous dialogue and the present hypothesis. A passive mode is one in which each new piece of information is offered to the physician in a pre-determined order. The latter technique is often actually used by doctors, who call it a CPC (Clinical Pathological Conference); the facts of the case are pre-arranged (often in a misleading manner) and read to a doctor who, at each stage, offers his or her current hypotheses and the reasons behind them. CPC's, unfortunately, are artificial in that the data is organized in ways which are foreign to a real doctor-patient interaction and the ensuing process may be unrepresentative of a doctor's normal strategy in making diagnoses. Thus, I have chosen to use a variation of the active process in which all the data about the patient is immediately available if the physician asks for it. This avoids assigning risks and costs to various diagnostic procedures, simplifying the problem to some extent. In this thesis, I will concentrate on the hypothesis-generation and evaluation aspects of the diagnostic process. I will not consider the question-asking strategy

in detail, except as it illuminates the more general topics of data organization and hypothesis generation. The protocol below (Chapter 2) was taken from a session in which the physician actively acquired data from the patient, although I have not included an analysis of the question-selection process in my work. The data structure arrived at in this thesis, however, should be amenable to the superposition of a question-selection module. Several strategies for asking questions are explored in <Gorry 74>.

1.3 The Basic Approach

Putting aside practical issues, one could formulate the diagnosis problem in terms of a classical maximum-likelihood schema: we have a collection of symptoms and a collection of diseases; the problem in each case is to choose the disease which is most likely causing the particular symptoms observed. In more general terms, we have a collection of effects and a collection of causes; the task is to find the cause which most likely accounts for the effects present in each particular situation. Under certain assumptions (which I will discuss below), the solution is straightforward and represents an elementary example of the use of probabilities. With each (disease,symptom) pair is associated a number which represents the probability of a patient who has the disease exhibiting the symptom. For example, if 20% of all people suffering from the flu have aching

muscles, then the number associated with (flu, aching muscles) would be .2. Obviously, the number implicitly associated with (flu, no aching muscles) would be .8. Then making a diagnosis necessitates only multiplying all the probabilities associated with present and absent symptoms for each disease - and comparing the results. The disease with the highest associated product is the winner and claims the victim.

This method is obviously generalizable to any recognition problem for which enough correlation data are available - given a few conditions:

1. that the symptoms are independent, in the probabilistic sense and
2. that the diseases are mutually exclusive and exhaustive.

Obviously, neither of these is true in the medical diagnosis case; patients often have more than one disease and the presence of one symptom more often than not affects the probability of the occurrence of others. Both of these non-linearities can, theoretically, be handled in the probabilistic framework by considering all possible combinations of diseases and symptoms in recording and combining probabilities. By now, an important reason for rejecting the above-outlined <u>complete theory</u> should be obvious: the uncontrolled proliferation of hypotheses and associated probabilities and the explosion of computations necessary to choose the correct answer. Even if all the numbers necessary were available (which they're not), this situation could become computationally infeasible - and is

certainly cognitively impossible. It doesn't take very subtle intuition to judge that doctors are not maintaining up-to-date "scores" on every possible diagnosis. In addition, when this approach is combined with similar methods for choosing tests, the amount of processing necessary quickly gets out of hand.

So the simple Bayesian theory seems untenable; the next step is to search for ways to reduce the number of hypotheses actively entertained at any given time and to cut down the amount of computation necessary to keep the relative status of each hypothesis up-to-date. The emphasis of the coming chapters will be on two stages in the movement away from a complete but unrealistic theory toward a heuristic theory which seems to model more closely the processing which physicians probably use. A brief summary of those two notions follows.

1.3.1 Activation vs. Deactivation: the first cut-back

The first mechanism has to do with the selection of hypotheses for active consideration. The complete theory postulates all diseases as possibilities from the beginning, eliminating them as their associated probability products go to 0. An obvious way to have fewer active hypotheses is not to consider a disease until it is suggested by a relevant piece of data. This has the reassuring consequence that every current hypothesis has a reason for being

remembered - instead of just lacking a reason for being forgotten. The issues surrounding this switch in emphasis are closely related to the concepts of expectation and evidence, which are discussed in detail in Chapter 4.

1.3.2 Heuristics and Interaction: the second cut-back

Both the complete theory and the modification discussed above are uniform theories; that is, every disease and symptom is treated the same. Some of the most powerful methods for controlling the growth of the hypothesis space, however, are much more specialized and local. They reflect knowledge about the non-independence of symptoms and the amount of detail a doctor must collect pertaining to a particular symptom before using it as a reason for considering a hypothesis. Such local pieces of knowledge will be viewed as compiled information, as they are derivable by general principles from the primitive data base of disease/symptom probabilities, but are clearly more efficient and useful in their specialized form. Chapter 5 contains an inventory of such interactions between symptoms and the imperative information associated with them.

In order to keep the number of active hypotheses at a reasonable level, it is important in addition to stop considering those whose plausibility has reached a low level and to avoid adding new hypotheses on top of old ones which have not yet been discarded as

useless. Such methods are clearly heuristic - that is, they don't always do "the right thing" - since any hypothesis we eliminate on heuristic grounds may eventually turn out to be the correct one after all. But it seems that physicians (and, most likely, all of us) must do everything they can to keep their minds uncluttered and their short-term memories from overflowing. Later sections discuss in more detail the postulated structures of both short- and long-term memory and their correspondence with the theory proposed here.

1.4 Anticipations

Chapter 2 contains a protocol of a doctor-patient interaction which illustrates many of the processes described above. The doctor is an expert; thus, modeling his reasoning means modeling expertise and we can expect many examples of compiled heuristics and special techniques. Chapter 3 describes a representational structure which we have developed in looking at hematuria and the diseases in which it plays an important part; the explanation of this data structure more clearly identifies the objects and relationships in a basic medical data base. Chapter 4 discusses the issue of local evaluation of hypotheses, making a distinction between disease-centered information (expectations) and symptom-centered information (evidence) and speculating on the place of each in a doctor's developing expertise. Chapter 5 catalogues some of the interactions betwwen symptoms which

contradict any strictly linear theory of evaluation - and which exemplify the compiled information mentioned above. Chapter 6 continues the movement from local toward global strategies by explicitly considering the structure of both simple and complex hypotheses and a theory of coherence designed to provide a way of comparing competing hypotheses and choosing the most promising ones. Chapter 7 summarizes the preceding view of medical diagnosis as a hypothesis generation and testing problem and includes some tentative thoughts on learning and further research. The Appendix contains the data on hematuria which was collected during this research and which forms the basis for the protocol and other examples quoted in the discussions.

The thesis will thus be a necessarily incomplete but hopefully illuminating look at the structure of a small part of medical knowledge and some processes which use that knowledge, throwing some new light on some of the basic paradigms of AI, as well as on the problem of medical diagnosis.

Chapter 2 - The Protocol

The contents of this chapter are a protocol of a simulated doctor/patient interaction in which Dr. Stephen Pauker played the patient and Dr. Jerome Kassirer the doctor. Dr. Pauker had access to the patient's chart and history and only volunteered data that was contained there. Dr. Kassirer was allowed to ask questions and only received information he specifically requested; the protocol is thus an example of active data acquisition. Although the analysis in the chapters which follow does not purport to explain a doctor's question-asking strategy, I will include interesting lines of questioning which Dr. Kassirer followed, especially when they illuminate the current hypotheses he was entertaining. After each newly-added finding, there is a discussion of the processing which Dr. Kassirer must have performed and a formalization of that procedure in terms of the theory proposed here, as well as some more general statements about other possibilities which were rejected and possible generalizations from the specific instance. Many of these comments were gleaned from Dr. Kassirer several weeks after the actual protocol was taken and thus represent a commentary from a rather different point of view.

2.1 The Protocol As A Reflection of Thought Process

Much of the work in this thesis and particularly that in this chapter is based on the assumption that a protocol gleaned from an experimental situation is an accurate reflection of the doctor's underlying thought processes. In fact, that assumption is probably unwarranted and we should be aware in our analysis that other factors, most notably the experimental situation itself, contributed to the conversation. Although a complete discussion of the relationship between the protocol and the actual diagnostic process is beyond the scope of this thesis, the following suggests some dimensions along which that distinction might be made.

Part of the instructions to the doctor were to list his hypotheses after every new finding and to explain his reasons for including or disregarding relevant diseases. Often I pushed him with questions such as "What about a tumor?", thus forcing him to explain why he had not mentioned certain possibilities. We might call his mode of response the explanation mode, as it included commentary on the diagnostic process as well as the decisions themselves. The necessity to explain and respond to my questions may have influenced Dr. Kassirer to actively consider (and perhaps immediately reject) more diseases than he would have normally. Where, in a real clinical situation, he may have responded to the presence of symptoms A and B with a single working hypothesis which he knew from experience to be

most probable, the experimental situation pushed him toward verbalizing more possibilities, even if their probabilities were lower.

The fact that Dr. Kassirer could generate explanations for most of his decisions is evidence for the existence of commentary on his decision rules. Often, pieces of raw data which a doctor learns in medical school (essentially the probability of symptoms given a disease - see Chapter 4) are utilized as explanations for rules like "If a patient has hematuria and a family history of kidney disease, consider poly-cystic-kidney-disease." As information is "compiled" into more efficient formulations, as described in the following chapters, the original knowledge is retained as an explanation - and also for debugging purposes, should that piece of compiled knowledge prove inaccurate or inapplicable.

Further attempts to distinguish between explanation mode and normal diagnostic thinking should follow the lines suggested above. In particular, we should be on the lookout for compiled knowledge/commentary pairs and realize that while the protocol exhibits extensive use of explanations, this may be an artifact of the experimental situation. For the time being, however, I will disregard this distinction and just try to account for the behavior exhibited in the protocol which follows.

2.2 The Technical Level

The finding-descriptions used both in the protocol and in the following chapters are not what a doctor expects to hear from a patient. Patients' descriptions of their symptoms are usually imprecise and certainly not in medical terminology. For example, although the protocol which follows starts with a finding of HEMATURIA, in the actual simulation, the patient entered the office complaining of "funny-colored urine." The doctor must take a number of steps, exemplified below, to translate the patient's report into a description on the "technical level;" I shall call this process validation.

In order to reduce complexity, I have decided to limit my investigation to symptom-descriptions on the technical level: those descriptions a doctor would expect from another doctor. In addition, the promotion of a patient's description to a more acceptable medical description has turned out to be a process which can be done locally in the cases I have examined; that is, a doctor usually tries to validate a symptom without using knowledge about the diseases it might suggest or its relationship to other present symptoms.

2.3 An Example of Validation Techniques - Funny-Colored Urine

A frequent patient complaint is "funny-colored urine." Such a finding could be a description of many pathological states, among them blood in the urine. A doctor has several techniques at his or her disposal to disambiguate the patient's description.

2.3.1 Lab Tests

Certain laboratory-type tests are guaranteed to determine what the underlying finding is. In general, the tests to be done and conclusions to be reached relevant to a patient's presenting symptoms can be arranged in a structure similar to a flowchart or decision tree, as in Diagram 2-1. The flow of control in the upper part of the diagram should be obvious: if pyridium, porphyrins and melanin have been ruled out, a Hematest is done to determine whether or not there is blood material in the urine; if it comes out positive, the sample is examined under a microscope for red blood cells; if negative, the Ictotest for bile is done and so on. In the squares are substances which are in the urine and causing its funny color. Notice that the diagram is truly procedural in that the tests carried out and conclusions drawn from test results like plasma color are dependent upon results of previous tests. Straw-colored plasma only indicates myoglobin when the hematest has been positive, but no red blood cells have been found under a microscope. BEETS are included as the "last

resort" guess; eating large quanties of beets can cause discoloration of the urine and a doctor might hypothesize this situation if no other etiology is found. Of course, no doctor would conclude the coloring agent was beets without making sure the patient had eaten them recently.

Representing this knowledge in both procedural and declarative forms points out some basic differences between the two types of formalisms. Note that the procedural representation forces an ordering on the component parts: sometimes that ordering is necessary, but sometimes it is just an artifact of the representation. In the funny-colored urine case, for example, just seeing red blood cells under a microscope is sufficient to conclude the patient has hematuria - and is, in fact, usually the procedure conducted to determine whether or not a patient has hematuria, while the Hematest may not be done. Another ordering artifact of this procedural representation is the placement of tests for pyridium, porphyrins and melanin before the Hematest. Epistemologically, the outcomes of those three tests have no effect on the interpretation of Hematest results. A strictly declarative representation, on the other hand, would make all these interdependences clear, but would not make any ordering explicit. Diagram 2-2 shows the same information in terms of evidence.

Neither the procedural or the declarative representation expresses the fact that more than one malady may be causing the discoloration. For example, pyridium may be used to treat a urinary

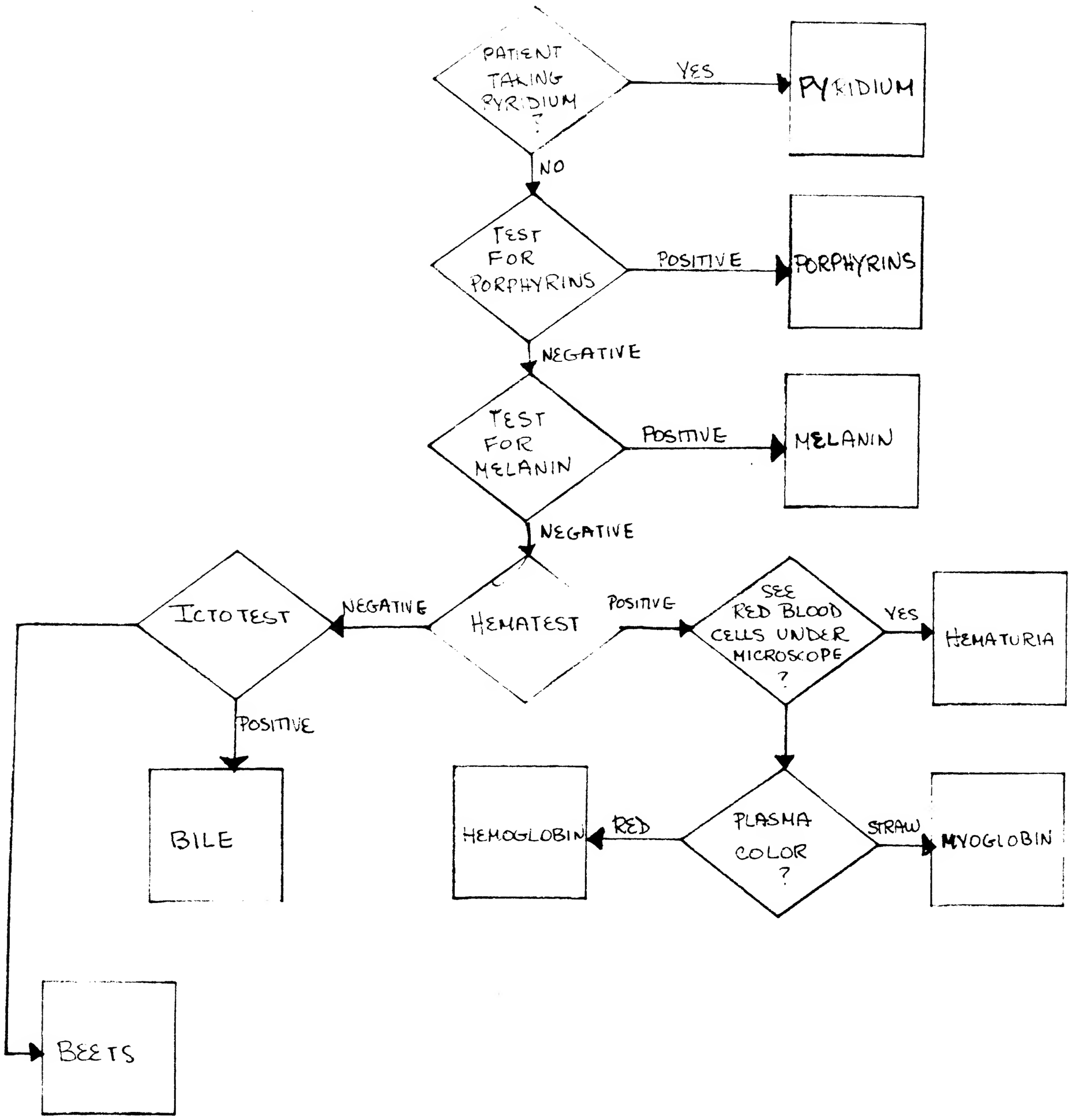


Diagram 2-1: Tests for Funny-colored Urine, Procedural Form

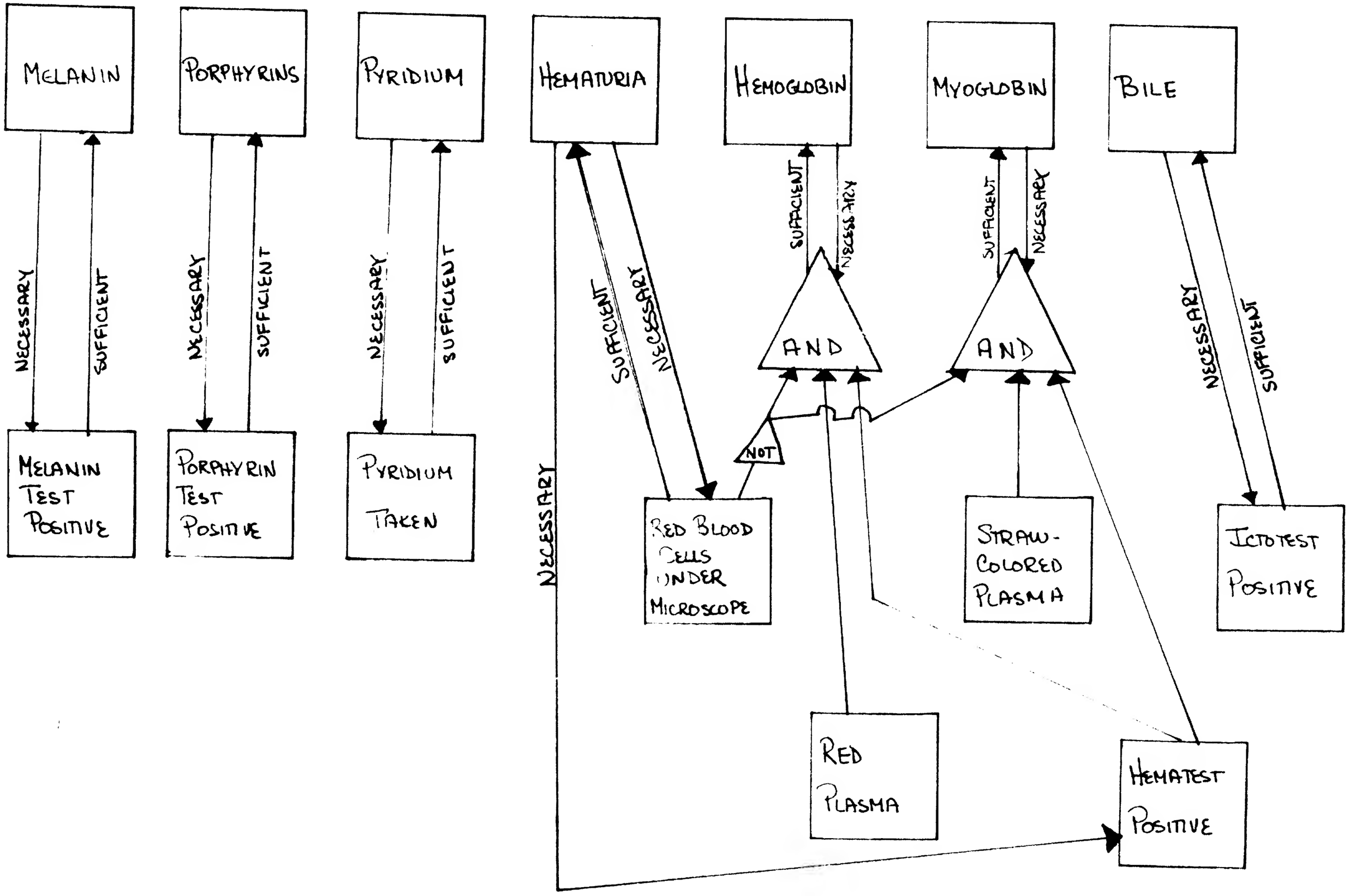


Diagram 2-2: Tests for Funny-colored Urine, Evidence Format (Declarative)

tract infection which is itself causing hematuria. In order to add this knowledge to the procedural form, we would have to add arrows from each of the squares containing coloring substances (e.g. pyridium) to the next test. Adding this same information to the declarative form would require specifying in the interpreter that all possible causes should be evaluated, even if another has already been confirmed.

A similar very local procedure exists for validating PEDAL EDEMA (fluid retention in tissues of the feet and lower legs) as the real problem behind the patient's complaint of puffy ankles. The doctor will usually press on the swollen area and observe how quickly and elastically the fluid fills up the depression; this procedure is carried out regardless of what other symptoms the patient exhibits.

2.3.2 Further Patient Data

Another technique which is used more often in the actual doctor/patient interaction is encouraging the patient to be more precise about his or her observations. Doctors ask questions in the patient's terms, not in medical terminology: in trying to pin down the funny color of urine, Dr. Kassirer often asks "Was it like cloudy tea? Coca-cola?" or in trying to determine the severity of a patient's shortness of breath, he may ask "How many flights of stairs can you climb? How many blocks do you walk from the bus stop home?" This

type of questioning is often used in conjunction with the lab tests mentioned above; if the urine color sounds characteristic of hematuria, only those tests relevant to blood in the urine would be performed.

2.3.3 Other Authorities

When the finding to be validated happened in the past, a doctor may have to resort to the opinions of other authorities. He or she may actually contact other doctors or, at least, ask the patient questions such as "Did any other doctor ever tell you that you had blood in your urine?" A doctor will also tend to interpret a past finding of funny-colored urine as hematuria if the present state of hematuria has been validated.

All of the above validation techniques are local in that they refer only to the finding in question, or other occurrences of the same finding at different times, not to possible diseases which could cause the finding or to other symptoms the patient might exhibit. A different approach would be to determine information about findings like PROTEINURIA; if proteinuria (protein in the urine) also existed, it would make hematuria more likely, as there is a disease which accounts for them both. Clearly, it is to a doctor's advantage to validate a particular finding locally, so as to cut down on the number of concurrent possibilities for the interpretation of a patient's

symptom description; if this is not always possible, other more global approaches may be necessary. This process obviously deserves much more study, as it must be carefully integrated into the diagnostic procedure which is the main topic of this thesis.

2.4 Competence vs. Performance

A question which often arises in the development of a theory or program is whether it is to represent the way a human would go about solving the problem - or, on the other hand, a procedure not subject to human failings like limited memory. The theory presented here leans strongly in the direction of modeling human processing and its major emphasis is on discovering the heuristics which doctors use in order to perform their task efficiently. Currently, theories of medical diagnosis are few and far between; many of the special heuristic measures presented here were discovered by watching real doctors, but may be necessary for <u>any</u> computational theory which can handle the vast amounts of medical information available.

However, every protocol is influenced by the diagnostic style of the physician and situational considerations; in a sense, our theory is still a model of <u>competence</u>, not <u>performance</u>. In going over the following protocol with Dr. Kassirer, I noted points where he could not explain his actions. For example, he actively considered PYELONEPHRITIS RECURRENT upon finding that HEMATURIA RECURRENT was a

symptom, but had not mentioned PYELONEPHRITIS as a possibility when confronted with just the symptom HEMATURIA. According to the theory proposed here, he should have at least initially considered the same diseases in both circumstances. We can explain such inconsistences by postulating that there are extraneous factors which affect the consideration of hypotheses; in many cases, one of those influences will be the limitations on a doctor's memory. Recent cases he or she has seen may come to mind more quickly, while others may be forgotten. The protocol itself, in addition, is not completely natural because at certain points the physician was pushed to make his hypotheses explicit; at those points, he may have mentioned a disease he had previously forgotten to mention, although he had considered it earlier.

We must be careful not to produce a theory which models too closely a particular doctor's behavior on one particular occasion, thus depriving it of its generality and power. Newell and Simon's <Newell and Simon 72> effort faces the same problem, as their data, like the data here, is taken from individual protocols. They comment, "Full particularity is the rule, not the exception. Thus, it becomes a problem to get back from this particularity to theories that describe a class of humans, or to processes and mechanisms that are general to all humans." (page 10)

2.5 The Informal Protocol

I will first present the case much as it happened, in English, with no discussion of the theory. The formal protocol which follows is, of course, a simplified and formalized form of the real patient/doctor interaction; both simplification and formalization are necessary in order to begin to develop a real theory of medical diagnosis.

The diseases which will figure heavily in this diagnosis are three glomerulitides: diseases which basically affect the glomerulus, a part of the nephron, the functional unit of the kidney. Glomerulitides are characterized by leakage of red blood cells and protein molecules into the urine, while they are usually trapped inside the blood vessels of the glomerulus. Acute glomerulonephritis (AGN) occurs several weeks after a streptococcal infection; it is probably caused by strep antigen-antibody structures affecting the glomerulus. Focal glomerulonephritis (FGN) is an episodic disease characterized by intermittent bouts of hematuria and proteinuria, separated by periods of complete normality. These episodes are often preceded by an upper respiratory infection a few days earlier. FGN may last the lifetime of the patient and doesn't seem to have any other bad effects. It also appears to be a form of familial nephritis - inheritable renal disease. Latent glomerulonephritis (LGN), on the other hand, is a progressive disease which eventually leads to chronic

glomerulonephritis (CGN) and renal failure. LGN is also characterized by stable proteinuria; that is, there is always some evidence of protein leakage into the urine.

The other disease which emerges as a possibility during the diagnosis is poly-cystic-kidney-disease(PCKD), a strongly hereditary malady which is evidenced by large cysts which form in the kidney, causing hematuria, high blood pressure and eventually renal failure.

The patient is a 31-year-old white woman who we will call Sarah. She was referred by another doctor to Dr. Kassirer, a nephrologist (kidney specialist). Sarah complains that her urine was funny-colored three days ago, but has been getting less dark since then.

Dr. K. Was your urine dark brown - about the color of Coca-Cola or cloudy tea when it was darkest?

Sarah Yes, that was the color.

Dr. K. Could I have the results of the Hematest performed today?

Sarah (Remember this is actually Dr. Pauker) Yes - there were 3 to 5 red cells per high power field.

Dr. K. And the rest of that urinalysis?

Sarah 1+ albumen and no red cell casts. (The former value is a quantitative measure of proteinuria made by dipstick.)

Dr. K. Have you had dark urine before? When was the last time?

Sarah A month ago. I've had intermittent dark urine for 10 years now.

Dr. K. Did you have any pain with the dark urine?

Sarah No, I didn't. Once or twice I had pain when I urinated and my urine was pink, but not with dark urine.

Dr. K. Those were probably unrelated urinary tract infections. Is there any history of kidney disease in your family?

Sarah My mother died of some kind of kidney disease when she was 40.

Dr. K. Is there any high blood pressure or deafness in your family? (Deafness is highly correlated with FGN)

Sarah No, but I've been taking medication for high blood pressure myself on and off for 5 years.

Dr. K. Has anyone in your family had a stroke? (Strokes are highly correlated with PCKD.)

Sarah No.

Dr. K. Let me get some lab results; what was the BUN? (an indicator of renal function)

Sarah It was 13 yesterday. (That's a normal value.)

Dr. K. I'll do the physical exam now: blood pressure 160/120; that's significant hypertension. Kidneys are not palpable. You've been coming to this clinic for several years, I see; what have the proteinuria measurements looked like?

Sarah At the last three visits, each six months apart, the 24-hour urine protein was 1650 mg., 480 mg., and 330 mg. (These are all slightly abnormal values.)

Dr. K. My diagnosis is that you have either LGN or FGN. LGN is a long-term disease which often lasts a long time - but in your case it has lasted unusually long, given the severity of your hematuria. FGN is a hereditary disease and there's a good possibility that's what you have, although I wouldn't have expected you to have proteinuria so consistently in a case of FGN. We should do a biopsy to decide between them. In any case, you have high blood pressure, so we'll treat that, but neither LGN nor FGN can be treated.

2.6 The Formal Protocol

Each finding in the formal protocol is first written in English, then in the formal representation explained below. The

current hypotheses follow, after which comes an explanation of the processing performed to generate and evaluate those hypotheses. I have tried to take as the current hypotheses those which Dr. Kassirer said he was entertaining during the session.

One place where the theory is relatively undeveloped is in designating exactly which hypotheses should be triggered or activated as the result of the addition of any particular finding. The more expert a doctor is, the more diseases he or she knows about; in order to handle so many possibilities, an expert's triggering process must be precise, activating only a few hypotheses. For example, the first symptom in the protocol is HEMATURIA, blood in the urine. Given only that symptom and the age and sex of the patient, Dr. Kassirer considered only three hypotheses; several other diseases are suggested by HEMATURIA (e.g. G-U-TUMOR, PYELONEPHRITIS etc.), but they were rejected or never actively considered by the doctor. Some of the heuristics involved in the triggering process will be discussed separately in Chapter 5.

Findings are represented by a type, a main-concept and a collection of property-value pairs. For example, in

```
SYMPTOM
        HEMATURIA
        PRESENCE    PRESENT
```

SYMPTOM is the type, HEMATURIA the main concept and PRESENCE PRESENT the relevant property-value pair. Chapter 3 contains more details on the representation of findings and disease hypotheses. The medical

data the doctor used in making this diagnosis is contained in the Appendix. The abbreviation G-U is used in several places for GENITO-URINARY (as in G-U-TUMOR and G-U-TRACT-BLEEDING). Several kinds of scores are used throughout the protocol as indicators of the likelihood of a hypothesis being valid. The significance of each type of score is explained immediately after its introduction in the protocol.

ENTER DOCTOR AND PATIENT

The patient is a 31-year-old woman; her name is Sarah.

FINDING1: FACT
PATIENT
SEX FEMALE

FINDING2: FACT
PATIENT
AGE 31

Comment: The first two items of data a doctor finds out about a patient are invariably age and sex; no hypotheses are generated until a presenting symptom is also mentioned.

****

Sarah had gross hematuria (blood in the urine) three days ago. (Actually, her initial complaint was funny-colored urine; see discussion above)

```
FINDING3: SYMPTOM
                HEMATURIA
                PRESENCE  PRESENT
                SEVERITY GROSS
                TIME  (AGO (DAYS 3))

HYPOTHESES:                                              score
GLOMERULITIS1:  GLOMERULITIS (AGO (DAYS 3))               1
AGN ISA GLOMERULITIS                                      1
        TIME (AGO (DAYS 3))
LGN ISA GLOMERULITIS                                      1
FGN ISA GLOMERULITIS                                      1
        (ISA FINDING3 EPISODE)
```

Explanation:

FINDING3 triggers or activates the general hypothesis GLOMERULITIS and three of its examples (members of its CHOICE-SET, as defined in Chapter 3), focal glomerulonephritis (FGN), latent glomerulonephritis (LGN) and acute glomerulonephritis (AGN). The first line under HYPOTHESES makes explicit that the GLOMERULITIS hypothesis inherits the time-indicator from the symptom; this is a time-instantiation of the GLOMERULITIS hypothesis and the name GLOMERULITIS1 is generated for it. As explained further in Section 3.3.4, the time of a disease is specified when it is hypothesized as the cause for a symptom which has a time-designation. In this case, the hematuria occurred three days ago, so we hypothesize that glomerulitis was present three days ago. Later, in the protocol, different time-instantiations of the same hypothesis will be meshed into a larger hypothesis and their respective scores combined.

The composite hypotheses formed by the global assembling stage of processing are listed below GLOMERULITIS1; each composite hypothesis has two elementary hypotheses joined by an ISA relation. Details of these complex hypotheses are contained in Chapter 6. Their scores represent their relative degrees of likelihood at this stage of the game. A complete discussion of the scoring algorithm and the rationale behind it comprises Chapter 4. A score of 1 essentially indicates that there are no discrepancies between the actual data and the expected disease description.

Considering FGN, an EPISODIC-DISEASE, requires interpreting this incidence of HEMATURIA as an EPISODE and the assertion (ISA FINDING3 EPISODE) is generated. AGN inherits the time-specification from HEMATURIA, while LGN does not, because LGN is labelled a LONG-TERM-DISEASE. The system described here does not handle time in a general way; obviously, a complete system would need a description of the time-course of, for example, AGN, which has two distinct stages of different durations. The time manipulations described here are sufficient for this protocol but will not handle all cases.

It is important to notice that certain obvious diseases which cause HEMATURIA are <u>not</u> entertained at this point, for many of the heuristic mechanisms which act to limit the number of hypotheses show up here. A few examples of GLOMERULITIS are not activated: chronic glomerulonephritis (CGN) is not considered because it can be adjoined to the LGN hypothesis, as happens later after FINDING10. It is, in

addition, pointed to by a differential-diagnosis pointer from AGN, as explained in Chapter 5. Systemic lupus erythematosus is also not activated; this may be a true case of memory lapse on the part of the doctor, for I could find no reason for its absence. G-U-TUMOR is triggered but immediately rejected because its a priori probability in a 31-year-old woman is very low. RENAL-INFARCTION, the death of kidney tissue due to lack of oxygen, has a similar fate. A priori probabilities are discussed in Chapter 4. They will not be systematically included in the scoring of each disease, but will be mentioned when they affect the processing, as when a particularly low a priori probability causes a hypothesis to be rejected.

PYELONEPHRITIS, infection in the kidney pelvis, is not considered because it requires HEMATURIA and PAIN (LOCATION FLANK) in order to be activated. CLOTTING-DISORDER also requires a combination of two findings to be triggered - for example, PREGNANCY and HEMATURIA. POLYCYSTIC-KIDNEY-DISEASE (PCKD) requires another finding like FAMILY-HISTORY of NEPHRITIS (kidney disease). A discussion of these multiple triggers and their heuristic value appears in Chapter 5.

****

A lab test done today showed 3 to 5 red cells per high power field in Sarah's urine: microscopic hematuria. Microscopic hematuria is less severe than gross.

```
FINDING4: SYMPTOM
                HEMATURIA
                PRESENCE PRESENT
                SEVERITY MICROSCOPIC
                TIME NOW

HYPOTHESES:                                          score      composite-score
GLOMERULITIS1: GLOMERULITIS                                          1
        START-TIME (AGO (DAYS 3))                      1
        END-TIME  NOW                                  1
FGN ISA GLOMERULITIS                                                 1
        (ISA GLOMERULITIS1 EPISODE)
LGN ISA GLOMERULITIS                                                 1
AGN ISA GLOMERULITIS                                                 1
        START-TIME  (AGO (DAYS 3))
        END-TIME NOW
```

Explanation:

The finding of microscopic hematuria at the present time triggered GLOMERULITIS again, instantiated this time with the time-specification NOW. Part of the local evaluation of GLOMERULITIS takes into account the two occurrences and combines them into a locally coherent hypothesis which represents the fact that these two symptoms are indicative of one occurrence of GLOMERULITIS. GLOMERULITIS1 is modified to show that it started 3 days ago and its END-TIME is stated as NOW (notice however we do not really know the END-TIME until it happens; the hematuria could last for several more days.) This combination represents the clustering of symptoms which suggest the same disease at different times; the theory generates only those hypotheses which are locally coherent in interpreting the several instances of HEMATURIA as part of the same disease process, rather than also considering less highly-valued hypotheses which interpret

the two occurrences as indicative of two different diseases.

A complicating dimension has been added to the scores here; we need a mechanism to combine the scores of different time-instantiations of a disease hypothesis. I have chosen simply to average the scores at different points of time, thus arriving at a composite score; in this case the computation is very simple, as both time-instantiations of GLOMERULITIS have scores of 1. In the theory presented here, this combining process over time-instantiations always occurs on a level more general than a particular disease, e.g. GLOMERULITIS or G-U-TRACT-BLEEDING. Specific diseases which are connected to these categories by ISA links inherit the composite scores; LGN, AGN, and FGN thus also have scores of 1 at this point. Often there is more precise time information relevant to the disease itself; if so, this is reflected in its time-score, while the score it inherits from a more general category is referred to as its symptom score.

This complex system of scoring is generally unintuitive and unsatisfactory; it is necessitated by the fact that diseases and symptoms occur over time. Hopefully, the development of more general flexible time representations (see Chapter 3 for more discussion) and better approaches to the interaction of time and symptomatology will provide a much better alternative. For the time being, however, the reader is requested to bear with this somewhat strange system.

****

The urinalysis also showed 1+ proteinuria (protein in the urine); the associated severity term is LIGHT.

FINDING5: SYMPTOM
PROTEINURIA
PRESENCE PRESENT
SEVERITY LIGHT
TIME NOW

HYPOTHESES:

GLOMERULITIS1 and the associated FGN, LGN, and AGN hypotheses remain unchanged;
each of them can account for PROTEINURIA,
and there is no change in their scores.

NEPHROTIC-SYNDROME rejected

Explanation:

The development of the GLOMERULITIS hypothesis and its examples AGN, FGN and LGN follows the pattern already exemplified above. GLOMERULITIS would have been rejected if the gross hematuria had occurred concurrently with the light proteinuria rather than three days earlier (see data-network in Appendix.) We notice here a restriction on local evaluation of hypotheses - it must be time-sensitive. In hypotheses such as GLOMERULITIS where the symptom occurs concurrently with the disease, it is easy to decide which findings should be considered for every instantiation of the hypothesis. (see Chapter 3 for a discussion of time-dependent instantiations); in cases where the suggestive finding occurs before or after the actual time of the disease, like an elevated ASLO-TITER

which occurs 1 to 5 weeks after a strep-infection, a simple calculation suffices to decide which findings are relevant to any particular instantiation. Section 3.3.4 on Time discusses these general issues in more detail.

Specifically, there are two time-instantiations of the GLOMERULITIS hypothesis being evaluated here. The one which was 3 days ago has only gross hematuria as its relevant finding. The one whose time is NOW has microscopic hematuria and light proteinuria, and there is no interaction specified between those findings. Again, AGN, FGN, and LGN inherit the score of GLOMERULITIS - none of them provides extra information in interpreting and scoring the findings.

NEPHROTIC-SYNDROME is negatively activated (see Chapter 4); it is ruled out without ever explicitly being considered a possibility. The NEPHROTIC-SYNDROME hypothesis has a NECESSARY EXPECTATION of heavy proteinuria,(3-4+ protein in the urine), which is violated by the finding of light proteinuria. Although there are certainly other diseases which wouldn't possibly fit the current symptoms, Dr. Kassirer explicitly mentioned the fact that NEPHROTIC-SYNDROME was ruled out.

****

Today's urinalysis also revealed no red-blood-cell casts.

FINDING6: SYMPTOM
RED-BLOOD-CELL-CASTS

```
                                PRESENCE ABSENT

HYPOTHESES:                                          score     composite-
                                                                 score
GLOMERULITIS1:  GLOMERULITIS                                     .75
        START-TIME (AGO (DAYS 3))                      1
        END-TIME NOW                                   .5
FGN ISA GLOMERULITIS                                             .75
        (ISA GLOMERULITIS1 EPISODE)
LGN ISA GLOMERULITIS                                             .75
AGN ISA GLOMERULITIS                                             .75
        START-TIME (AGO (DAYS 3))
        END-TIME NOW
```

Explanation:

RED-BLOOD-CELL-CASTS (PRESENCE PRESENT) is a MODERATE EXPECTATION in GLOMERULITIS. FINDING6 contradicts that expectation, thus making GLOMERULITIS NOW less likely, as its score of .5 indicates. This is the first time we come across any discrepancy between expectation and actual fact. The composite score for GLOMERULITIS, being the average of the two time-instantiation scores, also drops below 1. As before, FGN, LGN, and AGN's scores are simply inherited from GLOMERULITIS.

Gross hematuria is a possible excuse for the lack of RED-BLOOD-CELL-CASTS (see Chapter 5 for a description of excuses), but an excuse and the condition for which it is an excuse must be concurrent and in this case, the GROSS HEMATURIA occurred three days before the finding of no casts.

****

At this point, the doctor asked a lot of questions about the time course of the patient's hematuria, a strategy which culminated in his obtaining the information in FINDING7. It is interesting that he claimed he did not ask these questions specifically to differentiate between several current hypotheses. In fact, at this point in the protocol, Dr. Kassirer was much less explicit in his designation of hypotheses than I have been here. His strategy was, instead, symptom-specific; the questions are important ones to ask about HEMATURIA regardless of the hypotheses currently being entertained. This is one example of a local compilation of global information, a concept which is described in more detail in Chapters 4 and 5.

Sarah reported having had recurrent dark urine over the past ten years.

```
FINDING7:  SYMPTOM
                HEMATURIA
                PRESENCE PRESENT
                SEVERITY GROSS
                RECURRENCE RECURRENT
                TIME-RANGE (YEARS 10)

HYPOTHESES:                                                  score    composite-
                                                                      score

G-U-T-Bl:GENITO-URINARY-TRACT-BLEEDING                                 .92
        RECURRENT (TIME-RANGE (YEARS 10))          1
        (AGO (DAYS 3))                             1
        END-TIME NOW                              .75

GENITO-URINARY-TUMOR
        rejected because its TIME-INDEX contains
        ((DURATION (GREATER-THAN (YEARS 5))) VERY-RARE)

KIDNEY-STONE RECURRENT                          symptom-score    time-score
        ISA G-U-TRACT-BLEEDING
```

```
        (DURATION (YEARS 10))                    .92          .25

PYELONEPHRITIS RECURRENT
        ISA G-U-TRACT-BLEEDING
        (DURATION (YEARS 10))                    .92          .5

                                                 score        composite-
                                                                score
GLOMERULITIS1:GLOMERULITIS                                      .83
        RECURRENT                                 1
        (AGO (DAYS 3))                            1
        END-TIME NOW                              .5

                                                 symptom-score  time-score
LGN ISA GLOMERULITIS
        (DURATION (YEARS 10))                        .83           .25

FGN ISA GLOMERULITIS
        (DURATION (YEARS 10))                        .83            1
        (ISA GLOMERULITIS1 EPISODE)
        (ARE FINDING7 EPISODES)

AGN ISA GLOMERULITIS
        rejected because its TIME-INDEX contains (RECURRENCE NEVER)
```

Explanation:

We find here the most complex use of scores. As indicated above, the composite-scores are calculated by determining a separate score for each occurrence of the general disease category, such as GLOMERULITIS and G-U-TRACT-BLEEDING; these are then simply averaged together. The RECURRENT symptom is counted as one occurrence. The specific diseases inherit the composite-score of their category, while their time-scores are derived from their TIME-INDEX. Notice the symptom-scores of the hypotheses are very close at this point, but their time-scores are radically different.

There are three ways to interpret a RECURRENT SYMPTOM. The

first is to consider each recurrence an EPISODE in an EPISODIC DISEASE like FGN. When interpreting a RECURRENT SYMPTOM in this way, the system generates assertions like (ARE FINDING7 EPISODES) and (ISA GLOMERULITIS1 EPISODE), since an EPISODIC-DISEASE requires treating every occurrence of the symptoms as an EPISODE. The separate time-score indicates the likelihood of the disease's recurring for the amount of time indicated by the recurring symptom.

A second way of interpreting a RECURRENT SYMPTOM is to consider it suggestive of a RECURRENT disease for which the SYMPTOM is evidence. In terms of process, this involves using the symptom as a trigger and then checking to see if the TIME-RANGE on the SYMPTOM fits the RECURRENCE information on the hypothesis. This time, HEMATURIA GROSS triggers more possibilities in Kassirer's mind than it did before; I do not intend to try to explain this discrepancy in his performance, since it seems to be attributable to factors outside the scope of this thesis - possibly memory limitations and quirks. (See Section 2.3 above.) KIDNEY-STONE RECURRENT and PYELONEPHRITIS RECURRENT are both possibilities. As indicated above, the KIDNEY-STONE and PYELONEPHRITIS hypotheses are both examples of G-U-TRACT-BLEEDING and inherit their symptom scores from it. The score for G-U-TRACT-BLEEDING is less than 1 because there is only microscopic hematuria now, rather than gross. PYELONEPHRITIS is more likely to recur than KIDNEY-STONES, as the time-score indicates.

Finally, a RECURRENT SYMPTOM may be an indication of a disease

which is neither EPISODIC or RECURRENT; HEMATURIA GROSS can be intermittent in G-U-TUMOR, but this possibility is rejected because its TIME-INDEX contains the information ((DURATION (GREATER-THAN (YEARS 10)) VERY-RARE). Notice, also, that G-U-TUMOR had earlier been considered but rejected because of low a priori probability in a 31-year-old woman. Its re-appearance here is suggestive of a system, the details of which I have not worked out, in which DEFERRED hypotheses (see Chapter 3) may be marked with the reason for which they were deemed unworthy. If more compelling supportive evidence comes up, such a hypothesis may be reconsidered (and perhaps again rejected, as in this case). LGN, already a hypothesized etiology, can have HEMATURIA RECURRENT GROSS. Notice, however, that although the TIME-INDEX for the entire disease contains the information ((DURATION (BETWEEN (YEARS 0) (YEARS 10))) OFTEN), the particular symptom HEMATURIA GROSS is less likely to be present for such a long time. This piece of information affects the time-score of the LGN hypothesis, while the symptom-score reflects only the presence of HEMATURIA GROSS as a supportive piece of evidence.

It is also important to note that there are a lot of assumptions going into even the choosing of hypotheses to evaluate. The HEMATURIA RECURRENT could have been caused by, say, FGN, but the present episode be an indication of a KIDNEY-STONE. There are obviously some large number of such hypotheses which combine two or more explanations for the HEMATURIA. Doctors tend not to consider

them, though, unless forced to; they would rather think about the coherent hypotheses (see Chapter 6) which result from interpreting all the HEMATURIA episodes as indicative of the same etiology.

Up until now, I have been explicit about the hierarchical structure of the hypotheses and the evaluation of each hypothesis with respect to different times. Some of that detail is missing below, since it gets repetitive and boring; it should be remembered, however, that those more complex structures still exist explicitly in the system's representation of its current hypotheses.

****

Sarah reports having no flank pain associated with her hematuria.

```
FINDING8:  SYMPTOM
                PAIN
                PRESENCE ABSENT
                LOCATION FLANK
                RECURRENCE RECURRENT
                TIME-RANGE (YEARS 10)
                TIME-CONTEXT  (CONCURRENT-WITH  (FINDING7 FINDING1))

HYPOTHESES:                                      symptom-score    time-score
PYELONEPHRITIS RECURRENT      rejected
KIDNEY-STONE RECURRENT        rejected
LGN (DURATION (YEARS 10))                            .83              .25
FGN (DURATION (YEARS 10))                            .83                1
```

Explanation:

Both PYELONEPHRITIS RECURRENT and KIDNEY-STONE RECURRENT are rejected

because they expect PAIN FLANK; their symptom-scores become so low

because of this violated expectation that they are no longer actively considered. This lack of flank pain was concurrent with the HEMATURIA RECURRENT, as well as with the current episode. Thus the constraint on concurrency of symptoms in local evaluation is met. The score for the past episodes of both PYELONEPHRITIS and KIDNEY-STONE is 0 (+1 for HEMATURIA GROSS RECURRENT and -1 for PAIN FLANK ABSENT), so the hypotheses are rejected immediately.

LGN and FGN are unaffected, since the finding PAIN FLANK is not relevant to either of them. (see Chapter 4 for a definition of relevant symptom)

****

The doctor inquires about Sarah's family; his professed reason for doing this is because FGN is often a hereditary disease. Sarah says her mother had nephritis (a general word for kidney disease.)

FINDING9: FAMILY-HISTORY
NEPHRITIS
FAMILY-MEMBER MOTHER

HYPOTHESES:	symptom-score	time-score
LGN (DURATION (YEARS 10))	.83	.25
FAMILY-HISTORY NEPHRITIS	FACT	
FGN (DURATION (YEARS 10))	.875	1
POLY-CYSTIC-KIDNEY-DISEASE (DURATION (YEARS 10))	.58	1

Explanation:

The situation becomes much more complicated at this point. Up until the introduction of this finding, every hypothesis proposed was an adequate one, that is, it accounted for all the abnormal findings. We now find that one of the current hypotheses cannot account for the new piece of information. In this situation, we can either throw away the old hypothesis as inadequate or keep it around and add the new information as an independent finding. In general, the decision is a hard one, for patients often have more than one disease; I have a few suggestions, however, for principles on which to base the choice.

We are most concerned with accounting for abnormalities in the patient; accounting for FACTS and FAMILY-HISTORY findings is less important and they can be added to hypotheses as independent parts without much worry. (See Section 3.2.3 on Findings). Thus, in this case, we are allowed to complicate the LGN hypothesis. It is transformed, instead, into an LGN-centered hypothesis which has two independent parts, each with its own score. Later in the protocol are more examples of complicated hypotheses.

FGN can account for the FAMILY-HISTORY finding, so its score is actually increased.

In addition, a new hypothesis is triggered: POLY-CYSTIC-KIDNEY-DISEASE. Dr. Kassirer claimed that this disease has a multiple trigger - HEMATURIA and FAMILY-HISTORY NEPHRITIS - and thus wasn't activated before, although it is a common cause of

hematuria; Chapter 5 contains a discussion of multiple triggers and compiling. PCKD can account for all the symptoms, so it is an adequate hypothesis.

****

Dr. Kassirer asked Sarah whether or not she had ever had high blood pressure. He specified to the audience that the information would serve as a differential-diagnosis between PCKD and FGN, the first of which expects HYPERTENSION, while the second does not. Sarah reported having taken anti-hypertensive medication for 10 years.

```
FINDING10:  FACT
                ANTIHYPERTENSIVE-DRUGS
                STATUS TAKEN
                DURATION (YEARS 5)

HYPOTHESES:                                          symptom-score   time-score

LGN    (DURATION (YEARS 10))                            .83             .25
DEVELOPS-INTO
CGN     (DURATION (YEARS 5))                            .85              1
FAMILY-HISTORY NEPHRITIS                                 FACT

POLY-CYSTIC-KIDNEY-DISEASE
        (DURATION (YEARS 10))                            .72             1

LGN (DURATION (YEARS 10))                                .83             .25
FAMILY-HISTORY NEPHRITIS                                 FACT
HYPERTENSION ESSENTIAL
        (DURATION (YEARS 5))                             1               1

FGN (DURATION (YEARS 10))                                .875            1
HYPERTENSION ESSENTIAL
        (DURATION (YEARS 5))                             1               1
```

Explanation:

FINDING10 triggers HYPERTENSION CHRONIC and, in fact, is SUFFICIENT EVIDENCE for it, so HYPERTENSION CHRONIC is accepted. HYPERTENSION ESSENTIAL is also triggered, as accepted hypotheses which are not ULTIMATE-ETIOLOGIES act as findings in triggering possible CAUSES for themselves. (Because at this point HYPERTENSION CHRONIC is acting as a finding, rather than a hypothesis, it is not listed under HYPOTHESES.) Four coherent, adequate hypotheses are formed by the global assembling stage (see Chapter 6). This is another example of a finding which current hypotheses cannot account for, as HYPERTENSION is not relevant to either LGN or FGN. The recommendations here for incorporating such findings are, as stated above, merely a beginning.

The first hypothesis is LGN-centered and consists of LGN and CGN connected by the link DEVELOPS-INTO, as well as the independent FAMILY-HISTORY. The second is also LGN-centered but its third part is HYPERTENSION ESSENTIAL. The reason we are allowed to add HYPERTENSION ESSENTIAL as an independent part of the LGN-centered hypotheses is that its a priori probability for Sarah's age group is OFTEN. The PCKD hypothesis remains unscathed, as it can account for HYPERTENSION CHRONIC. A more complete system would have information about the relative times and durations of the various long-term symptoms (HEMATURIA and HYPERTENSION), but I have not included such knowledge. Finally, the FGN hypothesis is also complicated by the addition of HYPERTENSION ESSENTIAL, which is allowable for the same reasons of a

priori probability as in the LGN-centered hypothesis.

****

Some lab test results are available; the blood urea nitrogen (BUN), a major indicator of kidney function, is normal.

FINDING11: LAB-DATA
BUN
RESULT NORMAL

HYPOTHESES:	symptom-score	time-score
LGN (DURATION (YEARS 10)) FAMILY-HISTORY NEPHRITIS HYPERTENSION ESSENTIAL (DURATION (YEARS 5))	.83 FACT 1	.25 1
FGN (DURATION (YEARS 10)) HYPERTENSION ESSENTIAL (DURATION (YEARS 5))	.875 1	1 1
POLY-CYSTIC-KIDNEY-DISEASE (DURATION (YEARS 10))	.21	1
LGN (DURATION (YEARS 10)) DEVELOPS-INTO CGN (DURATION (YEARS 5)) FAMILY-HISTORY NEPHRITIS	is rejected	

Explanation:

The PCKD hypothesis has some expectation of RENAL-FAILURE CHRONIC and thus of an elevated BUN; this violated expectation, however, is not sufficient to reject the hypothesis. Chapter 6 explains the propagation of evidence from the RENAL-FAILURE CHRONIC hypothesis to the PCKD hypothesis. Actually, there is a time-dependence between the

onsets of hematuria, hypertension and renal failure in PCKD, and this knowledge would affect the evaluation of the hypothesis, but I have not represented it. There is a phase of PCKD where renal function has not yet begun to deteriorate, but where hypertension has already become a symptom. I have evaluated the lack of renal failure with respect to all the time-points for which PCKD has been instantiated, but additional time information would limit those evaluations to some subset of those times.

The FGN-centered and LGN-centered hypotheses which include HYPERTENSION ESSENTIAL are not affected by this finding, since BUN level is not relevant to any of the components of those hypotheses.

The LGN-CGN hypothesis has graver problems; RENAL-FAILURE CHRONIC is a NECESSARY EXPECTATION in CGN and BUN (RESULT HIGH) is a NECESSARY EXPECTATION in RENAL-FAILURE CHRONIC. Thus, the CGN component of the hypothesis is rejected. The whole hypothesis is then rejected because of a general assumption that each part of the hypothesis was necessary to account for some piece of data. In this case, CGN was added to account for HYPERTENSION. The way in which the connection is made between CGN and BUN level through a series of EVIDENCE links is detailed in Chapter 6 on global assembling.

****

During the physical-examination, the doctor discovers that Sarah does not have palpable kidneys (that is, he can not feel them from the outside).

FINDING12: PHYSICAL-EXAM
PALPABLE-KIDNEYS
PRESENCE ABSENT

HYPOTHESES:	symptom-score	time-score
PCKD is rejected		
FGN (DURATION (YEARS 10))	.875	1
HYPERTENSION ESSENTIAL (DURATION (YEARS 5))	1	1
LGN (DURATION (YEARS 10))	.83	.25
FAMILY-HISTORY NEPHRITIS	FACT	
HYPERTENSION ESSENTIAL (DURATION (YEARS 5))	1	1

Explanation:

PALPABLE-KIDNEYS are STRONGLY EXPECTED in PCKD and their absence makes that diagnosis so unlikely that it is rejected. Actually, PALPABLE-KIDNEYS are a NECESSARY EXPECTATION if PCKD has progressed as far as the duration of hypertension and hematuria would suggest, but again, I have not developed the facilities for dealing with this information. The other two hypotheses are unaffected.

****

Upon examining the patient's history more closely, the doctor found she had had slightly abnormal proteinuria every time she had been

examined over the past 10 years.

FINDING13: SYMPTOM
PROTEINURIA
SEVERITY LIGHT
DURATION (YEARS 10)

HYPOTHESES:	symptom-score	time-score
FGN (DURATION (YEARS 10))	.875	1
HYPERTENSION ESSENTIAL (DURATION (YEARS 5))	1	1
PROTEINURIA (DURATION (YEARS 10))	0	
LGN (DURATION (YEARS 10))	.83	.25
FAMILY-HISTORY NEPHRITIS	FACT	
HYPERTENSION ESSENTIAL (DURATION (YEARS 5))	1	1

Explanation:

This is the final symptom. Before its introduction, the doctor was convinced that the central diagnosis was either FGN or LGN; he was leaning strongly toward FGN because LGN seldom lasts ten years without turning into CGN and because the FGN-centered hypothesis could also account for the FAMILY-HISTORY. This final symptom, however - stable proteinuria over the past 10 years - is more representative of LGN. Since FGN is an EPISODIC-DISEASE, it would expect PROTEINURIA RECURRENT rather than stable. By the theory outlined here, we should really reject the FGN-centered hypothesis since it can't account for the PROTEINURIA and there is no coherent way to extend it which would account for this final finding. At this point, however, Dr. Kassirer couldn't decide between the two hypotheses and asked Dr. Pauker for

the pathologist's report on the biopsy - always the deciding diagnostic factor in a case like this. The same biopsy had been interpreted twice - once indicating FGN and once indicating LGN; it seems there is no clearcut diagnosis in this case. The distinction between the two diseases, however, is unimportant as far as treatment is concerned - neither responds to any known treatment. The difference lies mainly in the courses they will take; FGN will just continue benignly, while LGN will eventually develop into CGN and end-stage renal disease, which is often fatal.

The processes exemplified above will be explicated in detail in the following chapters. In particular, Chapter 3 details the data structure which underlies all the processes. Chapter 4 describes how each local hypothesis is evaluated, yielding the scores used in this chapter. Chapter 5 describes triggering and evaluation which take into account two or more symptoms, including excuses. It also talks about some distinctions between an interpretive theory and a compiled version of it which is presumably more representative of an expert's way of doing diagnosis. Finally, Chapter 6 will offer some principles for combining local hypotheses into coherent global ones, using chains of EVIDENCE pointers, CAUSE, COMPLICATION, DEVELOPS-INTO and ISA links. Chapter 7 contains some discussion of the real meaning of score. One point made there which is worth noting now is that the

precision implied by the numerical [illegible] chapter is misleading - it is an artifact of using numbers at all. Chapter 7 contains some first thoughts on other ways to look at the hypothesis-evaluation problem.

Chapter 3 - Basic Concepts of the Theory

This chapter will present the basic data objects and concepts of the theory, as well as preliminary details about process which are necessary to understand the next chapter. The data structure will be considered most generally as a net consisting of nodes representing causes and effects. The chapter concentrates on the detailed structure of nodes and on various relationships between symptom specifications in the data base and those presented in a particular patient. Finally, the processing states that certain nodes can find themselves in will be specified and a general overview of the evaluation procedure proposed for this data structure presented.

3.1 Cause, Effect and Mechanism

The knowledge which is necessary to do medical diagnosis has to do primarily with cause and effect. The data structure which is described in detail below realizes each cause or effect as a node in a knowledge net. I have called a node which is primarily an explanation or cause an elementary hypothesis. Those nodes which are basically not causes, but rather raw data, are called findings; several types of findings will be detailed later. Elementary hypotheses are differentiated from findings both by their ability to account for

(explain) one or several findings and procedurally by the fact that they are subject to a local evaluation process which determines how likely they are to be the diagnosis in light of the current data. There is really no clear line, however, especially since elementary hypotheses can themselves be effects which are explained by a more inclusive cause. SODIUM-RETENTION, for example, is the cause of a wide variety of symptoms, including WEIGHT (RANGE HIGH) and EDEMA; it, in turn, can be caused by AGN. Alternatively, a subset of findings called facts can sometimes be causes, as in a CATHETER causing HEMATURIA. Thus, the distinction I am making is somewhat artificial and not clear-cut. In general, elementary hypotheses represent diseases or pathological states whose existence must be inferred from findings which are in turn data more immediately obtainable from a lab test, physical exam or patient report.

A philosophical note on cause and effect: medical knowledge is not yet sophisticated enough to allow an analysis of disease analogous to a repairman searching for bugs in an electronic device. Medical researchers do not yet understand the human body well enough to be able to follow chains of cause-effect pairs back to the original malfunction. On a gross level, we can assert that FLU CAUSES UPSET-STOMACH or, in the renal world, PYELONEPHRITIS CAUSES CALYCEAL-CLUBBING. However, the mechanisms of those causations which, if understood, might yield some generalization which would make medical reasoning easier and more precise, are far from fully

explored. In essence, we don't yet know the structure of some of the body's basic circuits. Thus, we should expect a different kind of analysis from that suggested for electronics <Brown 74>. The procedure there is to use the description of each component's expected performance in the context of the circuit to localize the failure to a particular component. In medicine, such analysis is impossible; "cause and effect" fades into "correlated with" and the theory we come up with is one of hypothesis generation and testing. Even "testing" a hypothesis in the medical setting isn't as clearcut as in electronic troubleshooting, where a hypothesis can be easily tested by replacing a part and observing the circuit's behavior.

Of course, there are some instances in which the actual physiological mechanisms of a disease are known and it is interesting to speculate on the effect this knowledge might have on the diagnostic process. In fact, the functioning of the kidney is understood better than that of most other organs. For example, the route by which blood and protein molecules end up in the urine in glomerulitis is at least partially understood, as are the symptoms of sodium retention and its origins in glomerulitis (although the interactions of several proposed mechanisms for sodium retention have not been clarified.)

Knowledge of underlying mechanisms is clearly not necessary for doing diagnosis; knowing symptom-disease correlations is sufficient. Few doctors really understand the countercurrent system by which urine is concentrated, for example, but this lack doesn't

seriously affect their diagnostic skill. Knowing details of mechanism, however, does affect memory structure and thus the ease with which relevant symptoms are remembered. Although no current theory of memory is sufficiently detailed to explain the phenomenon, facts (in particular symptom/disease correlations) seem to be more easily remembered when accompanied by explanations. Perhaps the increased number of connections between the two concepts accounts for their easy recall; perhaps it is more profitable to think of the difference in terms of an increased number of access paths to the fact. In any event, facts for which a doctor knows an explanation can be regenerated if they are forgotten. A physician may, for example, forget whether URINE-SODIUM LOW or URINE-SODIUM HIGH is a symptom of GLOMERULITIS. By remembering that SODIUM-RETENTION is associated with GLOMERULITIS and realizing that, physiologically, increased sodium in the blood means less in the urine, he or she can rederive the correct symptom, URINE-SODIUM LOW.

In addition, knowledge of mechanisms is important in explaining diagnostic decisions - both to other doctors and to patients. This is, in addition, a relevant issue in considering the design of a computer program for diagnosis, since it must be able to explain itself to doctors who will be regarding it skeptically.

Studying the effect of such physiological explanations on memory structure will have to await better theories of both memory and medical diagnosis, but is an issue which may prove to be a valuable

pursuit in the future.

3.2 The Basic Components

3.2.1 History - Past and Present

Much of the data structure described here was influenced by a program written by Steve Pauker, in conjunction with William Schwartz and Tony Gorry. <Gorry 74> The project undertaken by Jerome Kassirer, Gerry Sussman and myself to examine the structure of medical knowledge surrounding hematuria, the presence of blood in the urine, grew directly from an examination of that program, which has also influenced many of the concepts presented later in this thesis.

Currently, Gorry et al are implementing a language called GOBBLE for representing and retrieving medical knowledge. Their system addresses directly some of the representation issues referred to here.

The data represented here and in the Appendix is incomplete in two ways : much of the actual data, such as relevant symptoms and the time course of diseases, is not included. In addition, the relative amounts of evidence which different symptoms contribute to various hypotheses has not yet been specified. My purpose in examining the data structure only to this extent was to identify basic important concepts and structures enough to explore the processes described in later chapters. Our efforts have also provided a lot of data for an

eventual system which will know "everything" an expert nephrologist knows about the differential diagnosis of hematuria; although the details must be filled in, the skeleton of the knowledge base is already worked out.

3.2.2 Elementary Hypotheses

Each cause node representing a disease, syndrome or pathophysiological state is called an elementary hypothesis. A number of findings which are correlated with the disease are associated with each elementary hypothesis; they are called relevant findings. We call the entire structure of elementary hypothesis and associated findings a slice; the connections between findings and hypotheses within a slice are intraslice connections. Most of the findings in a disease's slice are abnormalities which are caused by the disease, but facts about a patient's age, sex, race or family history may also be included in the slice. (For more discussion, see Chapter 4)

Take, for example, the two slices represented in Diagram 3-1, URINARY-TRACT-INFECTION and PYELONEPHRITIS. Each of them is an elementary hypothesis. The findings associated with URINARY-TRACT-INFECTION (UTI) include FEVER, HEMATURIA, FREQUENCY etc. Those associated with PYELONEPHRITIS include PUS-CASTS, PAIN (LOCATION FLANK) and IVP (RESULT SCARRING) There are several other elementary hypotheses on the page: GLOMERULITIS and IRRITATION are included to

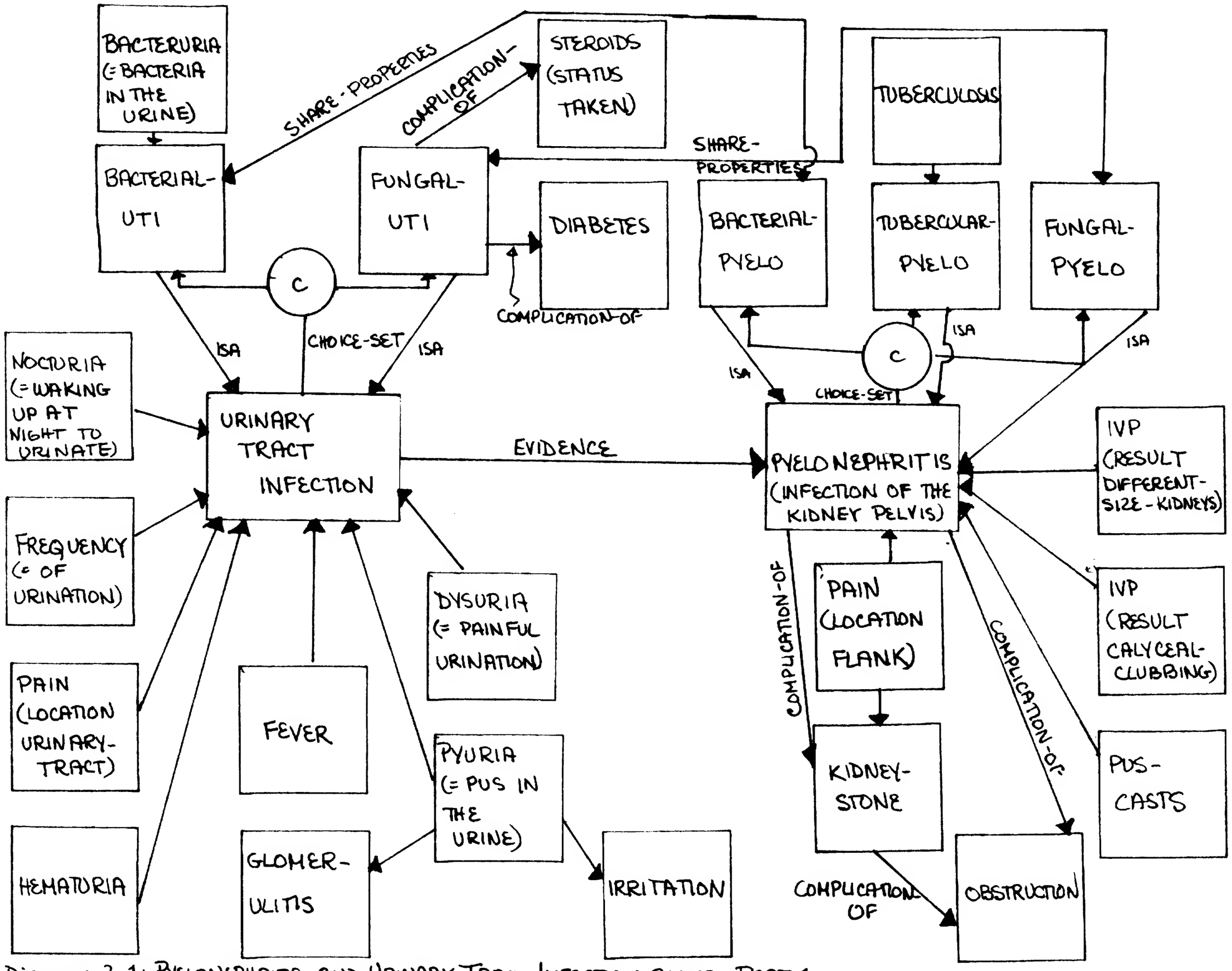


Diagram 3-1: PYELONEPHRITIS AND URINARY TRACT INFECTION SLICES: PART 1.

PYELONEPHRITIS	EVIDENCE/EXPECTATION VALUES
RELEVANT-SYMPTOM	
PUS-CASTS	
PRESENT	+SUFFICIENT
IVP	
CALYCEAL-CLUBBING	+SUFFICIENT
PAIN	
FLANK	+STRONG
ABSENT	-STRONG

TIME-INDEX:

(RECURRENCE (BETWEEN (YEARS 0.) (YEARS 20.)) SOMETIMES)

(SUFFICIENT-CHOOSER BACTERURIA BACTERIAL-UTI in UTI)

(SUFFICIENT-CHOOSER TUBERCULOSIS TUBERCULAR-PYELO in PYELONEPHRITIS)

Diagram 3-1a: PYELONEPHRITIS AND URINARY TRACT INFECTION SLICES: PART 2

illustrate that a single finding may exist in many slices. In the data structure depicted here and in the Appendix, both elementary hypotheses and findings are contained in rectangles or squares; there is no distinction between squares and rectangles; I use both only to fit the whole slice on one page. The symptoms of a disease (relevant findings) are connected to its elementary hypothesis by pointers which have been left unlabelled. Because EVIDENCE is the most common relationship between nodes, only the other more unusual ones have been explicitly marked. The designations of EVIDENCE and EXPECTATION strengths, whose derivation is explained in Chapter 4, are included below the diagram or on the following page, where there was insufficient space, in those cases where they have been determined. I have sometimes included short definitions of the medical terms inside their rectangles; they are included in parentheses and preceded by an "=" to differentiate them from property-value pairs.

An elementary hypothesis may be regarded as a structure which helps to organize data. In current psychological theory, the concept of "chunking" has become popular as a way to explain various memory phenomena. Briefly, Short-Term Memory (STM) is assumed to contain a small (7 + or - 2) number of places, each of which can hold one "chunk" of information. In the case of a doctor trying to make a diagnosis, we can consider each place occupied by a finding or elementary hypothesis. Clearly, if several pieces of data are chunked into a single hypothesis, they will take up only one space in STM. If

a patient has DYSURIA, HEMATURIA and IVP (RESULT SCARRING) the doctor can organize that knowledge into a hypothesis about PYELONEPHRITIS. In one simulated case in which Dr. J. P. Kassirer was asked to make a diagnosis, however, the facts fell into no single hypothesis and he was forced to write them down to remember them. The symptoms in that case were: COMA, HYPERTENSION, ANEMIA, CATARACTS, INFECTION (a hypothesis itself subsuming FEVER and WHITE-BLOOD-CELL-COUNT HIGH) and RENAL-DISEASE (a hypothesis supported by PROTEINURIA HEAVY). Dr. Kassirer's question-asking and hypothesis-generation style in this case was much less directed and efficient than normal because the data he had were not organized into a single hypothesis.

3.2.2.1 Properties of Elementary Hypotheses

Often diseases have general properties which aid in their diagnosis. Three which I have singled out are time-related properties, EPISODIC-DISEASE, ABRUPT-ONSET-DISEASE and LOMG-TERM-DISEASE. They are indicated in the data diagrams by thin rectangles attached to the bottoms of elementary hypotheses; see Diagram 3-2 for an example. These properties are "distributive" in that they really describe the findings associated with the diseases; the way this distribution is handled is dealt with in section 3.3.1 on Fitting. RENAL-INFARCTION, the death of renal tissue due to interference with the circulation, is an ABRUPT-ONSET-DISEASE, meaning

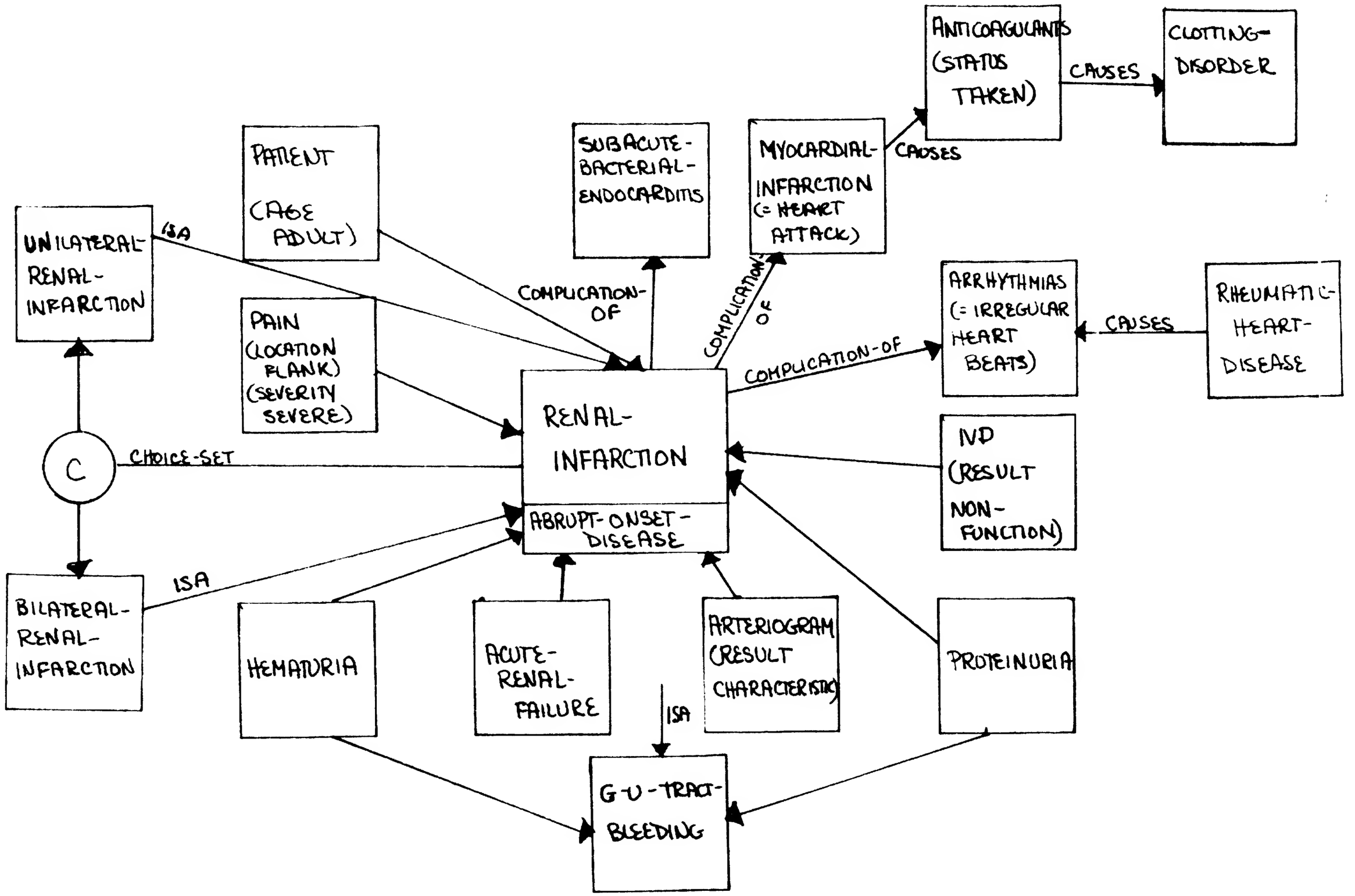


Diagram 3-2: RENAL-INFARCTION SLICE

each of its symptoms appears quickly. We expect, for example, that PAIN SEVERE will start suddenly, at approximately the same time as HEMATURIA does. Focal glomerulonephritis (FGN) is an EPISODIC-DISEASE, that is, it consists of several episodes of HEMATURIA separated by periods of no HEMATURIA. Latent glomerulonephritis, on the other hand, is a LONG-TERM-DISEASE which lasts many years, although it is not episodic. The beginnings of a method to handle these properties is contained in section 3.3.4 on Time.

In addition, elementary hypotheses may or may not be ULTIMATE-ETIOLOGIES. An elementary hypothesis which is an ULTIMATE-ETIOLOGY is one which could stand alone as a diagnosis, for which a more basic cause does not have to be sought or is not known. Given the current state of medical knowledge, it is sufficiently specified to be a diagnosis and to recommend particular treatment. For example, GLOMERULITIS is not an ULTIMATE-ETIOLOGY, although FGN, LGN, and AGN are. HYPERTENSION is not an ULTIMATE-ETIOLOGY; HYPERTENSION ESSENTIAL is, as it is HYPERTENSION considered to have no identifiable cause. Similarly, NEPHROTIC-SYNDROME is not an ULTIMATE-ETIOLOGY, although it may be included in an overall hypothesis as a COMPLICATION-OF GLOMERULITIS. IDIOPATHIC-NEPHROTIC-SYNDROME, on the other hand, is an ULTIMATE-ETIOLOGY which can stand alone.

Each elementary hypothesis also has an associated TIME-INDEX

relating to its expected duration and recurrence. These are described in more detail in Section 3.3.4 on Time.

3.2.2.2 Relations Between Elementary Hypotheses

I have described the symptom/disease data structure as a network of causes and effects. Elementary hypotheses can sometimes themselves be considered symptoms, so that EVIDENCE and EXPECTATION pointers may connect them. In Diagram 3-2, UTI and PYELONEPHRITIS, both elementary hypotheses, are so related. I allow the EVIDENCE/EXPECTATION pair of links between two elementary hypotheses only if, in general, the symptoms of the disease entity which is being considered analogous to a symptom are also symptoms of its associated cause. (The symptoms which are exceptions may be pointed out by OVERRIDE assertions, as explained in Chapter 5.) This is indeed the case with UTI and PYELONEPHRITIS, as well as with SODIUM-RETENTION and AGN. Such chains of symptoms (e.g. EDEMA is a symptom of SODIUM-RETENTION which is a symptom of AGN) are the result of the grouping of symptoms of a disease into subgroups which have a single mechanism and thus also occur together in other diseases.

There are also clearcut CAUSE relations between elementary hypotheses where the symptoms of the two diseases concerned are not in a subset/superset relation. In this case, the CAUSE relation is stated explicitly, as in STREP-INFECTION CAUSES AGN. Similar to CAUSE

links are COMPLICATION links, as in PYELONEPHRITIS is a COMPLICATION of STONE. There seems to be a subtle medical difference between these two concepts: CAUSE represents a situation where the mechanism of causation is known, while COMPLICATION is more a designation of "closely correlated," in which the mechanism is not quite as clear. As far as process is concerned, they are treated identically by the system.

Closely related to these two properties is DEVELOPS-INTO, a relation which encompasses both time and symptomatology. In general, the relationship between the symptoms relevant to one disease and those relevant to one it DEVELOPS-INTO is again a subset/superset relationship, the direction dependent on whether the disease generally gets better or worse. For example, LGN DEVELOPS-INTO CGN; the symptoms of CGN are all those of LGN plus HYPERTENSION and RENAL-FAILURE. AGN1 (the active phase of AGN) DEVELOPS-INTO AGN2 after a few days; the symptoms of AGN2 are only a <u>subset</u> of those of AGN1 - namely HEMATURIA and PROTEINURIA - since it represents an improvement in the condition of the patient. Since I am unsure of the generality of this result, all the symptoms for each of the diseases (or stages) in a progression will be explicitly stated in the data structure.

Some elementary hypotheses are examples of others - more specific designations of etiology. STREP-PHARYNGITIS, SCARLET-FEVER and STREP-SKIN-INFECTION are all examples of STREP-INFECTION. The

links I use to designate these relationships are ISA and CHOICE-SET. The ISA link goes from the more specific example to the general category; the CHOICE-SET of a category is its set of examples. Acute glomerulonephritis (AGN), focal glomerulonephritis (FGN) and latent glomerulonephritis (LGN) are all examples of GLOMERULITIS; therefore, AGN ISA GLOMERULITIS, FGN ISA GLOMERULITIS and LGN ISA GLOMERULITIS and the CHOICE-SET of GLOMERULITIS contains (FGN LGN AGN) as well as some other diseases. A CHOICE-SET is indicated in the diagrams by a circular node marked "c"; from it come pointers to all members of the CHOICE-SET. If a CHOICE-SET is EXHAUSTIVE, it is so marked in the diagram; otherwise, no assumption is made. Because GLOMERULITIS is also used as a name for a collection of symptoms (namely HEMATURIA, PROTEINURIA and RED-BLOOD-CELL-CASTS), it is also joined to the members of its CHOICE-SET by EVIDENCE pointers. A category may have more than one CHOICE-SET. G-U-TUMOR has CHOICE-SETS corresponding to both location choices (KIDNEY vs. BLADDER vs. URETER etc.) and malignancy (BENIGN vs. MALIGNANT).

A final connection between elementary hypotheses illustrated in the data structure is SHARE-PROPERTIES. The idea is really similar to the variable-binding and matching mechanism implemented in PLANNER and CONNIVER, but it has been singled out explicitly in the data structure here. SHARE-PROPERTIES essentially enforces the equivalence of two variables. In the examples I have used, it shows up as a relation between members of two CHOICE-SETS, or a symptom and a

CHOICE-SET, but the concept of sharing information between different structures is clearly a more general issue and is really orthogonal to the CHOICE-SET concept. Examples should clarify this idea. The CHOICE-SET of PYELONEPHRITIS is BACTERIAL-PYELO, FUNGAL-PYELO and TB-PYELO, while that of URINARY-TRACT-INFECTION (UTI) is BACTERIAL-UTI and FUNGAL-UTI. In addition, UTI is evidence for PYELONEPHRITIS. More specifically, though, FUNGAL-UTI is evidence for FUNGAL-PYELO and BACTERIAL-UTI for BACTERIAL-PYELO. The SHARE-PROPERTIES pointer requires that when a choice is made in either CHOICE-SET, it is checked for consistency against the other entity. As in PLANNER and CONNIVER, if the other choice has already been made, it must agree; if not, the appropriate member of the second CHOICE-SET must be chosen. Another example is the BENIGN/MALIGNANT CHOICE-SET of G-U-TUMOR. The malignancy of a BIOPSY used as evidence must be the same as the malignancy of the hypothesized TUMOR; if the TUMOR has not yet been marked for malignancy, the BIOPSY finding makes that choice.

3.2.3 Findings

The system knows about several different types of findings: LAB-DATA, PHYSICAL-EXAM, SYMPTOM, FACT and FAMILY-HISTORY. LAB-DATA, PHYSICAL-EXAM and SYMPTOM are treated equivalently and the differentiation is instead made for epistemological completeness and future expansion. In future considerations of doctors' strategies in

active acquisition of data, it will be necessary to take into consideration the normal order a patient/doctor interaction follows: history, physical exam, lab tests (except for urinalysis and other quick tests done before the doctor meets with the patient). LAB-DATA in particular may be difficult or dangerous to obtain; the theory proposed here, however, does not deal at all with cost/benefit analysis. SERUM-CREATININE (RANGE HIGH) is a LAB-DATA; PALPABLE-KIDNEYS is a PHYSICAL-EXAM and PAIN (LOCATION FLANK) is a SYMPTOM. FACTS and FAMILY-HISTORY are treated differently from the other three in the final global process (see Chapter 6), since they don't have to be accounted for or explained in the same way as symptoms - yet, they affect the final diagnosis significantly. PATIENT (AGE YOUNG-ADULT) and CATHETER (PRESENCE PRESENT) are FACTS while FAMILY-HISTORY NEPHRITIS is an example of FAMILY-HISTORY.

Aside from its type, a finding consists of a main-concept and several property value pairs. For example, in the finding LAB-DATA SERUM-CREATININE (RANGE HIGH), the main concept is SERUM-CREATININE and the value of the property RANGE is HIGH. The property name is usually redundant, since it is uniquely determined by the concept and the property value so I will often omit it. Another example: LAB-DATA THROAT-CULTURE (RESULT POSITIVE) (TYPE BETA-HEMOLYTIC), in which the concept is THROAT-CULTURE, POSITIVE is the value of RESULT and BETA-HEMOLYTIC is the value of the property TYPE. A property value may also be a negation, as in SERUM-CREATININE (RANGE (NOT HIGH)).

There is also a distinguished property PRESENCE and its associated values PRESENT and ABSENT. A finding must have at least one property-value pair; if there is no other relevant property, PRESENCE can always be used. Therefore, HYPERTENSION (PRESENCE PRESENT) is a legal finding or finding-specification, as is HYPERTENSION (PRESENCE ABSENT), but HYPERTENSION is not. In the text here, however, I will omit the designation PRESENT where it is redundant.

Both Steve Pauker's and Tony Gorry's programs contain dictionary routines which know how to determine property names from values and which list the properties and associated values which are allowed for each main-concept. A dictionary must also contain an indication of what the normal value is for each property, so that the final diagnosis can account for all abnormal findings. For most findings, ABSENT or NORMAL is an expected property value indicating a finding which does not have to be explained. Since the designation of LAB-DATA, PHYSICAL-EXAM etc. is usually irrelevant and otherwise obvious from the other elements of the finding, I will often leave it out in describing findings.

A concept specified by one or more property values bears the same relationship to the unmodified concept as a member of a CHOICE-SET does to the category governing it; there is, in a sense, an implicit ISA link from the modified concept to the unmodified one. They are both examples of the same descriptive mechanism, an insight which has been worked out most thoroughly within the context of frames

(see Chapter 7). There are two reasons why both modes of expressing the same subset/superset relation are available:

1. CHOICE-SETS apply primarily to elementary hypotheses, rather than findings. In those cases, it is important that the possibility of generating the less specific hypothesis exists, since often the information which distinguishes among the more specific diseases isn't available until later. For example, G-U-TUMOR is often suggested by symptoms such as HEMATURIA and WEIGHT LOW, before we have any idea about its location or malignancy.

2. In those instances where a CHOICE-SET has been used on a finding, as opposed to an elementary hypothesis, it is for the purpose of asserting its influence (via SHARE-PROPERTIES) on another CHOICE-SET. The relevant example here is the BENIGN/MALIGNANT choice on the BIOPSY finding, a choice which determines the same property in the TUMOR CHOICE-SET.

In addition, CHOICE-SET members and modified symptom nodes are treated differently during the diagnostic process. The modifiers on a symptom node are designations of property-values which must be filled before a patient symptom can be accepted. CHOICE-SET designations, on the other hand, are distinctions which are made after the activation of the more general category.

3.2.3.1 Other Relations Between Findings and Elementary Hypotheses

As mentioned above, sometimes a finding is a CAUSE for another finding or elementary hypothesis; some examples found in the data here are (ANTI-COAGULANTS TAKEN) CAUSES CLOTTING-DISORDER and CATHETER CAUSES TRAUMATIC-BLEEDING

3.3 Fitting Patient Facts Into the Specification-network

The process of deciding whether or not a particular patient symptom is relevant to the symptom description in the knowledge network and thus is relevant to the disease hypothesis is called fitting: it requires trying to fit a particular finding-description into a sometimes more general specification. This notion and terminology comes from frame theory, as does the idea of further specification (see Chapter 7). To continue some of the frame terminology, we call each finding description in the data network a slot which would like to be filled with an actual finding (an instance). The attempted fit can result in one of several outcomes, which are detailed in the following sections.

3.3.1 Sufficient or Further Specification

We only try to fit a symptom into a slot if its main-concept and type are the same as the slot's. If that is true, then a

comparison is made between corresponding property-values in the slot and in the patient data. If they are the same, the data obviously fits the slot; this circumstance is called sufficient specification. If, in addition, the data contains a value for a property not mentioned in the slot, it is a further specification of the slot and can fill it. Note that this means that the negations of properties must be explicitly stated in the slot-specification. For example, the data NOSE RUNNY RED will fit the slot NOSE RUNNY since it is a further specification; if we were interested in NOSES which were just RUNNY, we would have to specify NOSE RUNNY (NOT RED) in the slot description. Closer to home, EDEMA PITTING ERYTHEMATOUS would fit EDEMA PITTING but not EDEMA PITTING (NOT ERYTHEMATOUS). The obvious thing happens when data such as PROTEINURIA (GRAMS 3) attempts to fit into the slot PROTEINURIA (GREATER-THAN (GRAMS 2)); the logical relationship is used to determine whether or not the data fits; the same thing happens in the slot-description (EDEMA (OR MASSIVE PITTING)). When a finding fits a slot, some change will occur in the score of the hypothesis to which that slot is attached, as explained in Chapter 4.

3.3.2 Insufficient Specification

If the data does not contain a value for a property specified in the slot description, but matches it in all other respects, it is an insufficient specification and the data does not fit. Thus,

THROAT-CULTURE POSITIVE does not fit THROAT-CULTURE POSITIVE BETA-HEMOLYTIC. Faced with incomplete information such as this, a doctor often asks more questions to obtain a specific enough symptom on which to base his or her hypotheses. This question-asking strategy has been explored in more detail by Pauker, Gorry and Schwartz in their study of EDEMA <Gorry 74>. From an AI point of view this strategy can be regarded as a local effort to reduce the number of active hypotheses; the more specific a symptom, the fewer diseases it will be applicable to. For example, SORE-THROAT (SEVERITY SEVERE) may suggest, among other diagnoses, TONSILITIS, STREP-PHARYNGITIS and PHARYNGEAL-HERPES. Adding (APPEARANCE WHITE-SPOTS) to the symptom description makes the diagnosis almost surely PHARYNGEAL-HERPES.

An example of this question-asking strategy appeared in the protocol in Chapter 2, where Dr. Kassirer asked the patient a number of questions about the time course of her hematuria; this attempt to further ascertain the properties of the symptom was without reference to possible diagnoses; it was, instead, a local procedure which is essentially compiled from global knowledge about what information would differentiate between various diseases. The general concept of local compilation of global knowledge is a thread which extends through this entire thesis; Chapter 4 examines its implications in more detail, while the protocol in Chapter 2 and the examples of interactions in Chapter 5 provide more specific instances of its realization and importance.

3.3.3 Contradictory Specification

A finding-specification is <u>contradictorily specified</u> if a piece of patient data makes its presence impossible. The absence of the data is then considered a violated expectation and the corresponding hypothesis has its score which represents its likelihood of being present in the patient lowered. Contradictory specification can happen in several ways. In cases where the values for a particular property are mutually-exclusive, a finding which has a different value for a property than the slot-specification is a contradictory specification. For example, THROAT-CULTURE POSITIVE ALPHA-HEMOLYTIC is a contradictory specification to THROAT-CULTURE POSITIVE BETA-HEMOLYTIC. An obvious example is HYPERTENSION PRESENT vs. HYPERTENSION ABSENT. In general, a PRESENT/ABSENT juxtaposition is only a contradictory specification if <u>all</u> the other values match. Thus, NOSE RUNNY ABSENT says nothing about NOSE RED PRESENT. Finally, a slot-specification such as NOSE RUNNY is contradictorily specified by a finding NOSE (NOT RUNNY).

We need to be able to pinpoint contradictory specifications because they represent discrepancies between expected symptoms and fact; this discrepancy is then reflected in the score of the associated elementary hypothesis. In view of the multitude of ways to

obtain contradictory specifications, at least one programmer (Steve Pauker) has resorted to spelling out everything explicitly. Thus, HYPERTENSION ABSENT might be an explicit slot-specification, in a disease where HYPERTENSION rarely showed up as a symptom, rather than just being omitted from that disease's slice. This is probably the way to go as far as implementation is concerned, but for conceptual ease, I prefer to be able to talk about EVIDENCE and EXPECTATION as presented in the next chapter. Even there, however, I will note that in cases of differing severities for the same symptoms, a more explicit data structure is necessary.

3.3.4 Time

A discussion of time must take into account two separate issues: how to represent the relevant information in the data diagrams and how to use that information in the interaction which occurs in diagnosis between the data structure and the patient's symptoms. Both aspects of the problem are complex and I will deal with each only incompletely.

Representing the time course of diseases requires expressions like:

(BEFORE AGN STREP-INFECTION (INTERVAL (WEEKS 2 3)))

which says that a related strep-infection precedes AGN by one to three weeks. Kahn has developed a competent system of time-indicators along

the lines of the expression above <Gorry 74>; his system understands the relations BEFORE, AFTER and DURING, can talk about the START and END of events or states and uses the abbreviations (AGO (WEEKS 5)) for (BEFORE NOW (WEEKS 5)). In addition, it has a fuzz-factor for all time measurements, reflecting the fact that in real life, the placement of events along a time line is often inaccurate; people generally divide their lives into childhood, high-school, college etc. and may not have events totally ordered within those subcategories. Kahn's system uses a fuzz-factor of, for example, several years in designating the time of a childhood disease:

(TIME-OF MEASLES (AGO (YEARS 40)) (FUZZ (YEARS 3)))

This approach to the representation of time is closer to the way people do it than placing all the events linearly along a time line with exact specifications of the distances between points; further investigations of people's internal representations of time will probably indicate an even more qualitative view of time, in which events are chunked into typical days, weeks, months and seasons, some of which have relative temporal orders, while others are unordered.

In the diagrams in the Appendix, time relationships of the BEFORE/AFTER type are indicated by an arrow marked BEFORE; the amount by which one state precedes the other dangles from that pointer.

In addition, there is more general information about the time courses of many diseases. I have mentioned above the designations EPISODIC-DISEASE, LONG-TERM-DISEASE and ABRUPT-ONSET-DISEASE. We will

see below how each of these properties affects the diagnostic procedure. In addition, each elementary hypothesis may have a TIME-INDEX which specifies facts about its DURATION and RECURRENCE. For example, G-U-TUMOR may have as part of its TIME-INDEX

(DURATION (GREATER-THAN (YEARS 5)) VERY-RARE)

The frequencies which fill the last place of such expressions are NEVER, VERY-RARE, RARE, SOMETIMES, OFTEN and ALWAYS which have corresponding values 0, 0, .25, .5, 1 and 1 for use in time-scores, which are explained below. VERY-RARE is included as a separate value, although it is treated the same as NEVER, as explained in Chapter 6. In addition, a particular symptom-specification may have some time-specification in it. As we saw in the protocol in Chapter 2, LGN has a general TIME-INDEX which contains:

(DURATION (BETWEEN (YEARS 10) (YEARS 15)) SOMETIMES)

but the following symptom is only WEAK EVIDENCE for the disease.

(HEMATURIA (SEVERITY GROSS) (RECURRENCE RECURRENT)

(TIME-RANGE (GREATER-THAN (YEARS 5)))

Microscopic hematuria for ten years is a common occurrence in LGN, but not gross hematuria for that long. In this case, the specific information about the symptom overrides the more general information about the disease.

What happens when a particular finding is added to the data base for a patient? How does it interact with the representation described above? General time-properties affect the process of

fitting a finding to a slot. If a disease is an ABRUPT-ONSET-DISEASE, the finding must contain the specification ABRUPT-ONSET in order to fit.

When a finding-specification is not connected to an elementary hypothesis by any time relationships, it is assumed to be concurrent both with the elementary hypothesis and with the other symptoms in the slice. In order to keep findings which occur at different times separate, I use the notion of a time-instantiation of an elementary hypothesis, an instance in time in which that disease is postulated to have been present, because of the presence of its symptoms. When an elementary hypothesis is evaluated, each of its time-instantiations is evaluated separately with its relevant symptoms; the scores are then averaged to produce a composite score which takes account of the times of all relevant findings. Thus, given the general piece of information

(BEFORE STREP-INFECTION (ASLO-TITER (RANGE HIGH))
(INTERVAL (WEEKS 1 5)))

(ASLO-TITER (RANGE HIGH) (TIME NOW)) would not be used in the evaluation of the elementary hypothesis (STREP-INFECTION (TIME NOW)), but would create another instantiation of STREP-INFECTION with (TIME (AGO (INTERVAL (WEEKS 1 5)))). I call these hypotheses which combine several occurrences of the same symptom locally coherent; this concept will be discussed further in Chapter 6. For example, in the protocol in Chapter 2 we had the following two findings:

(HEMATURIA (SEVERITY GROSS) (TIME (AGO (DAYS 3))))

(HEMATURIA (SEVERITY MICROSCOPIC) (TIME NOW))

The GLOMERULITIS hypothesis thus had two time-instantiations, one for each occurrence of HEMATURIA. These were then combined into one occurrence of GLOMERULITIS, whose START-TIME was (AGO (DAYS 3)) and whose END-TIME was NOW; its composite score was calculated as indicated earlier in this section.

The elementary hypotheses, then, are objects which do not have any absolute time of occurrence, although they may have temporal relationships to other elementary hypotheses and findings. Relating an actual patient symptom to the timeless elementary hypothesis instantiates it with an absolute time, "anchoring" this particular occurrence of the disease in time.

In the examples I have pursued, this method of scoring by averaging scores of individual time-instantiations to obtain a composite score has been most useful in the general hypotheses GLOMERULITIS and G-U-TRACT-BLEEDING where there is no specific information on the expected time-course of the pathological state. The more specific diseases like FGN, AGN and PCKD inherit the composite score of their category as their symptom score and in addition a time score is calculated for each disease which represents a more disease-specific interpretation of the time information. The TIME-INDEX of each disease contains the information necessary for this calculation.

Instead of having an absolute time designation, a finding may be RECURRENT; a RECURRENT finding may also have an associated TIME-RANGE and a TIME-CONTEXT which specifies its temporal relationship to other findings. (See FINDING8 in the protocol)

(HEMATURIA (SEVERITY GROSS) (RECURRENCE RECURRENT)
(TIME-RANGE (YEARS 10)))

is an example of such a recurrent symptom. When such a finding occurs, it can be interpreted in one of three ways. It may be considered evidence of an EPISODIC-DISEASE like FGN; if so, an assertion of the form (ARE EPISODES <finding>) is generated and the TIME-INDEX is consulted to determine how commonly the disease recurs for the length of time designated by the TIME-RANGE of the finding. Secondly, the finding which is RECURRENT may be considered an indication of a disease which is recurring, such as PYELONEPHRITIS RECURRENT; again the time-score is determined from the TIME-INDEX. The relevant data for PYELONEPHRITIS would be something like:

(RECURRENCE (BETWEEN (YEARS 5) (YEARS 10)) SOMETIMES)

Third, certain non-recurrent, non-episodic diseases may have recurrent symptoms; LGN is such a disease, as HEMATURIA often recurs over several years in LGN. In these cases, the DURATION part of the TIME-INDEX contains the relevant information for the time-score. In the protocol, G-U-TUMOR was considered as a cause for HEMATURIA RECURRENT (TIME-RANGE (YEARS 10)), but was rejected because its TIME-INDEX contained

(DURATION (GREATER-THAN (YEARS 5)) VERY-RARE)

The global assembling process described in Chapter 6 also makes use of time information. If it is trying to construct a coherent hypothesis, the temporal relationships in the instantiated elementary hypotheses must match those specified in the data diagrams. For example, a coherent hypothesis consisting of STREP-INFECTION and AGN would have to adhere to the specification:

(BEFORE STREP-INFECTION AGN (INTERVAL (WEEKS 2 3)))

While the above mechanisms handle many of the specific problems I encountered in my study of hematuria, this approach to time has a major problem. The scores of hypotheses are based primarily on the findings, not their temporal relationship; the symptoms are, in a sense, the primary consideration and time only secondary. Future diagnosis systems should be aware of this dichotomy and study it accordingly. Disease processes, however, occur over a period of time and it is often the pattern of a disease over time which clinches the diagnosis, rather than the symptoms at any one point in time. We can think of trying to map a description of the patient's state over time into a general description of a disease, sliding the two temporal descriptions along each other until the "best" match occurs; such a process makes the time-course of a disease the primary consideration. We should also be less cavalier about the designations RECURRENT and ABRUPT-ONSET, as a doctor (and thus this system) must be able to construct such descriptions out of more primitive data and reports of

individual occurrences of the symptom.

3.4 Overview of the Evaluation Process

3.4.1 States In Which Nodes May Be

During the course of a diagnostic session, nodes of the data network change state with the addition of new information about the patient. Findings have the fewest number of possible states, partly since I have purposely limited them in that way. A finding-specification may be either confirmed or disconfirmed if we have the relevant information or unknown if we do not. A finding-specification is confirmed by a sufficient or further specification; it is disconfirmed by a contradictory specification.

Of course, this strictly binary view of findings is not a true reflection of a doctor's data structure, as mentioned in the comments preceding the protocol in Chapter 2. Much of a good diagnostician's time goes into validating a patient's descriptions of his or her present state and past medical history, through tests, questions, contacting other authorities and looking up old records. A more detailed example of this validation process concerning funny-colored urine is contained in Chapter 2.

Elementary hypotheses, because they are not directly confirmable, have a more complicated set of alternative states. When a diagnostic session starts, all elementary hypotheses are inactive;

that is, no particular disease has been suggested by the patient's symptoms. As more data is presented, certain hypotheses become active by virtue of their correlation with and ability to account for the findings present; actually, as noted above, an active time-instantiation of the hypothesis is set up, as opposed to the elementary hypothesis itself. Once a hypothesis is active, it is evaluated after the addition of every finding to see how well it fits the data so far and some score is produced which represents the likelihood of that disease's being present. On the basis of this process, a hypothesis may be accepted or rejected; in most cases, however, no definite decision will be made, but its score will be modified to reflect the effect of the new data. An accepted elementary hypothesis is one for which the evidence is sufficiently specific to rule out any other cause for the symptoms present. For example, the presence of RED-BLOOD-CELL-CASTS confirms the diagnosis of GLOMERULITIS, making it an accepted hypothesis, but the very same finding makes SICKLE-CELL-TRAIT a rejected hypothesis. Elementary hypotheses may also be accepted or rejected when their scores reach certain threshold values; for more discussion on this point, see Chapter 4.

The final processing state which we can attribute to an elementary hypothesis is deferred. In the overall attempt to reduce the number of concurrently-active hypotheses, certain possibilities may not be considered active, even though they have been suggested by

a particular symptom. One basis for deferring a hypothesis is the a priori probability of the disease, especially given the age and sex of the patient. (see Section 4.3.3.2 for more on a priori probabilities.) For example, a doctor seeing a RASH on a patient's body may think of MEASLES or CHICKEN-POX; if the patient is a child, those hypotheses should certainly be followed up, but if he or she is an adult, they are deferred because they are so unlikely. The reason there is a distinction between deferred and rejected is that a deferred hypothesis can be resurrected at a later time by more symptoms which suggest it. In the example above, an adult could have MEASLES or CHICKEN-POX and if other symptoms supported either of those hypotheses, it would have to be considered more seriously. Deferred hypotheses should be marked with a reason for which they were rejected; the more serious the reason, the more evidence is necessary to re-activate the hypothesis. Although I have not worked out the details, it is clear that something like this is going on in a doctor's head.

3.4.2 Four Major Steps

How does the magic transformation from a bunch of symptoms to a final diagnosis take place? The process seems to be divided into four steps: disposing, triggering, local evaluation and global assembling. This series of four steps is performed after the addition

of each finding. If at some point the finding being added is designated as the last one, a diagnosis will be attempted; if not, another finding is added and the four stages performed again. I will give a brief synopsis of each stage here in order to make what follows more coherent; a top-level flowchart of the control structure is included as Figure 3-3; the data flow is detailed in Figure 3-4. Triggering and local evaluation are examined in more depth in Chapters 4 and 5 and global assembling in Chapter 6.

The major data structures used in the processing are the data network which contains the medical information and several lists which hold findings and hypotheses during the course of the diagnosis. The FINDING-LIST contains all the findings, each marked NORMAL or ABNORMAL. The ACTIVE-LIST contains all active elementary hypotheses; the ACCEPTED, REJECTED and DEFERRED LISTS contain the elementary hypotheses in the corresponding state. The ACCEPTED-LIST also contains those FACTS which can act as explanations, for use in the disposing phase. In addition, the COHERENT-HYPOTHESIS-LIST and the ADEQUATE-HYPOTHESIS-LIST contain hypotheses containing more than one elementary hypothesis which are built during the global assembling stage.

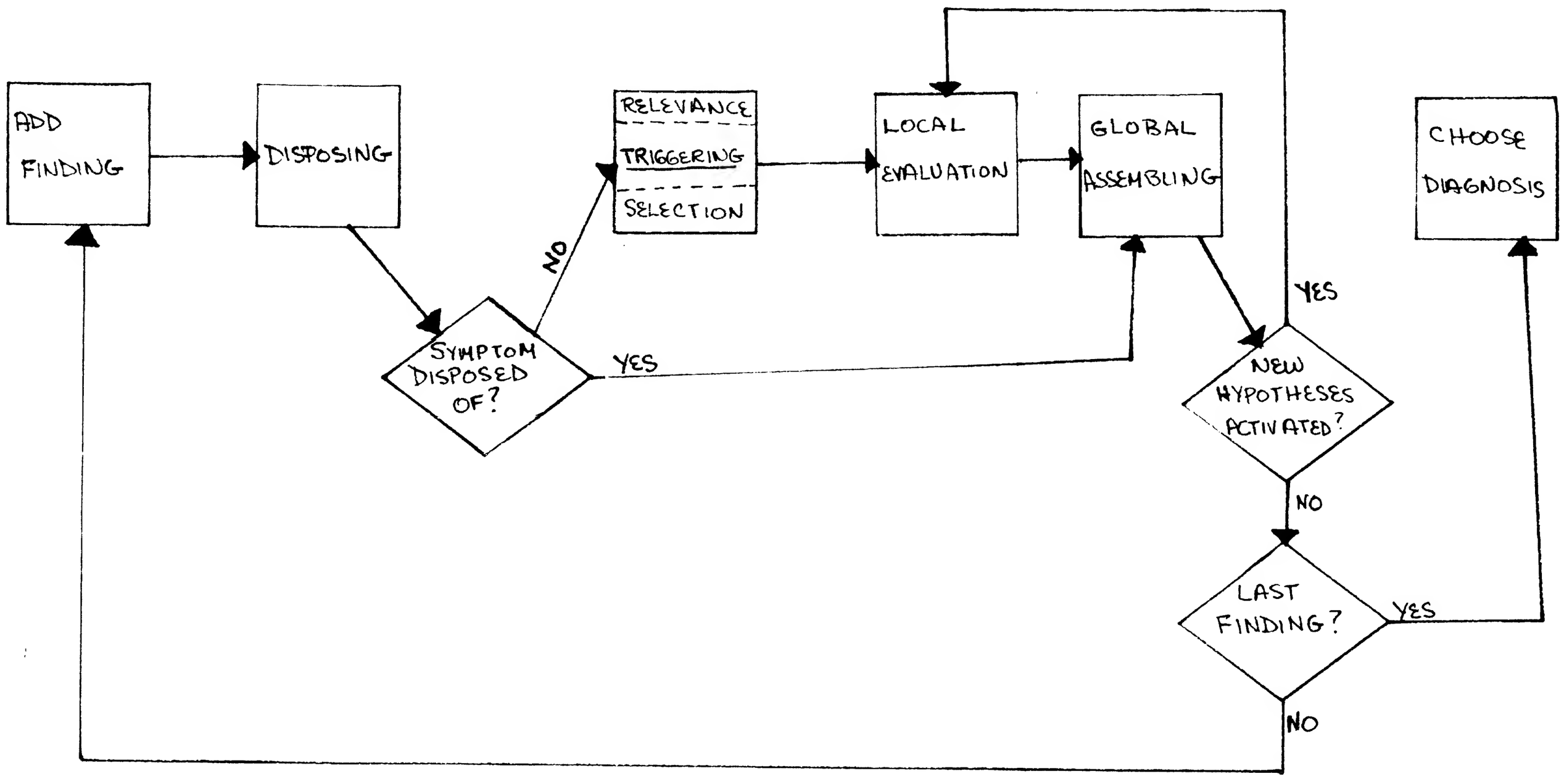


Diagram 3-3: CONTROL FLOW IN THE SYSTEM

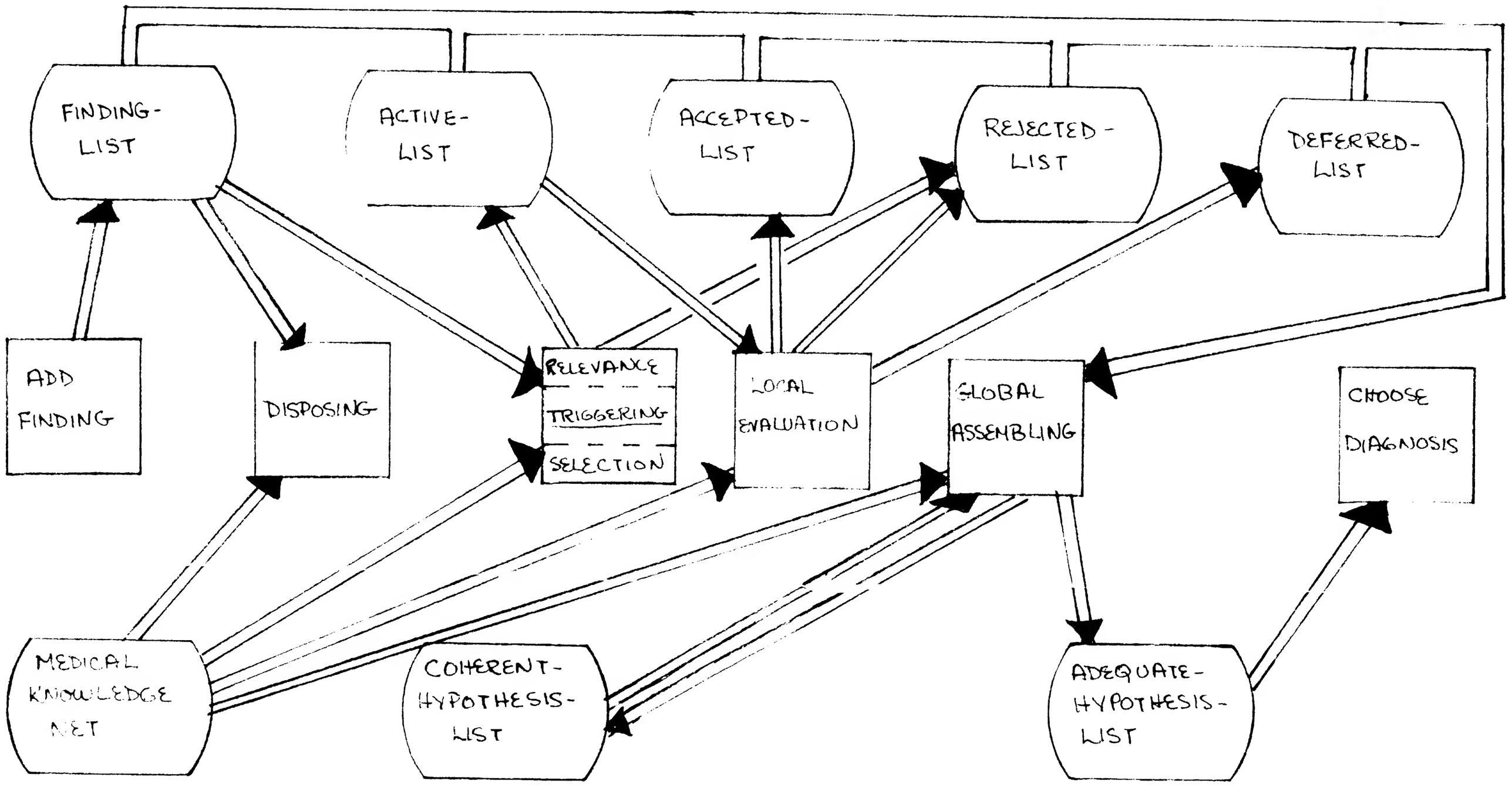


Diagram 3-4: DATA FLOW IN THE SYSTEM

3.4.3 Disposing

Sometimes the cause of a finding is clear when the finding is encountered; this is most often the case when the explanation is a FACT. In such a circumstance, the doctor does not bother to look for other explanations; although the symptom may have two concurrent causes, considering this possibility would mean greatly expanding the number of active hypotheses. Since I have argued many times above that an overabundance of hypotheses is to be avoided, it seems reasonable to try to dispose of a finding as a result of some already-accepted etiology rather than trying to find a new explanation. For example, suppose a patient is brought into the emergency room of a city hospital after an automobile accident; if his urine contains blood, the doctor should surely attribute it to abdominal trauma, rather than considering GLOMERULITIS. Similarly, a CATHETER PRESENT in a post-operative patient is often the cause of HEMATURIA. Most of these relationships are contained in the data graphs as explicit CAUSE relationships between findings; what I have called the disposing stage (not to be confused with garbage collection) searches for such a accepted explanation and if one is found, the triggering and local evaluation stages are skipped.

3.4.4 Triggering

Triggering is one of the processes by which an elementary hypothesis makes the transition from the inactive to the active state. A subset of the symptoms which are relevant to a disease are marked as triggers. When a symptom is asserted to be present in the current case, it activates all those elementary hypotheses for which it has been designated a trigger. For example, DYSURIA (painful urination) triggers URINARY-TRACT-INFECTION; NAUSEA by itself triggers nothing, as it is a common finding in many disorders. The activated hypotheses are added to the ACTIVE-LIST and the symptom itself is added to the FINDING-LIST. Elementary hypotheses may also be activated during the local evaluation and global assembling phases by mechanisms which will be dealt with in detail later.

Triggering, although at first glance a simple concept, has some of its own complexities. Conceptually, we can divide the process of choosing the right hypotheses to activate into two parts which Winograd <Winograd 72> has called relevance and selection. The relevance section consists of matching a finding to a trigger-specification by only a subset of its properties, perhaps just main-concept. Then more complex processing may take place to see if the symptom really fits and if the hypothesis is to be among the selected ones. Hairy pattern matchers often implicitly contain this division in their MATCH routines. The first step in matching is

finding candidates which fit a description like "anything with an A as the third element." Only then are more complicated checks like restrictions on the types of other elements carried out. Evidence of this two-step process comes in doctors' remarks such as : "HEMATURIA suggests STONES, but I wouldn't expect a stone to last 10 years." and "HEMATURIA suggests RENAL-INFARCTION, but the onset wasn't abrupt." The negative activation phenomenon to be discussed in 4.3.2 may be regarded as an example of this two-step process. In the protocol, after FINDING5, NEPHROTIC-SYNDROME was considered relevant because one of its triggers contains the main-concept PROTEINURIA; the attempt to fit the symptom into the slot, however, revealed a contradictory specification, so the hypothesis was rejected out of hand.

Sometimes elementary hypotheses are activated by a combination of two symptoms; POLY-CYSTIC-KIDNEY-DISEASE, for example, is activated by HEMATURIA and FAMILY-HISTORY NEPHRITIS, but not by either one alone. The selection part of the triggering process can also check for the presence of another symptom and select or disregard the proposed hypothesis accordingly. Triggers and multiple triggers are examples of local compilation of global knowledge; this will be explored more fully in Chapters 4 and 5.

3.4.5 Local Evaluation

As described in detail in Chapter 4, each elementary hypothesis has an associated local evaluation function which produces a value representative of how likely the disease is to be present given the data. Each of the hypotheses on the ACTIVE-LIST is evaluated, taking the new finding into account. In general, findings which are present add evidence to a hypothesis, while those which are expected but absent make it less likely. For some hypotheses, there may be no change in state, since the finding may not be relevant to the hypothesis. For others, however, drastic changes may occur: hypotheses may be accepted, rejected or deferred on the basis of the new finding. New diseases may, in fact, be suggested through the differential diagnosis mechanism explained in Chapter 5, added to the ACTIVE-LIST and evaluated in turn.

The local evaluation functions are basically linear, taking account of each symptom separately and independently. Sometimes, however, one symptom's presence or absence affects the significance of the other symptoms. Chapter 5 deals in particular with these non-linearities in local evaluation.

The evaluation done at this stage is local in that the functions do not ask questions about the status of other elementary hypotheses, or consider symptoms other than those relevant to the disease hypothesis being evaluated. These matters are left to the

fourth step, global assembling.

3.4.6 Global Assembling

The purpose of this fourth stage is to arrange the various local elementary hypotheses into a larger structure which fulfills two criteria: it is coherent and it is adequate. The rules of coherence have to do with the ways to connect various elementary hypotheses through links like CAUSE, COMPLICATION and EVIDENCE. Often this involves activating a previously-inactive hypothesis and then evaluating it, so there may be a cycle back to the local evaluation stage of processing. For example, if SODIUM-RETENTION and GLOMERULITIS are both active, the rules of global coherence allow us to active AGN1 and evaluate it. This stage of processing obviously uses the ACTIVE-LIST, the ACCEPTED-LIST and the connections inherent in the data network. An adequate hypothesis is one which accounts for all the abnormal findings in a case, as saved on the FINDING-LIST. Clearly, an adequate hypothesis is the end goal of a diagnostic process. An adequate hypothesis must also include every accepted elementary hypothesis. Sometimes forming such a hypothesis requires assuming that some symptoms are unrelated to others: that the patient has two or more unrelated diseases. This happened in the protocol, where the doctor was forced to assume the patient had HYPERTENSION ESSENTIAL, an etiology unrelated to the "chief" diagnosis of FGN. The

notion of ULTIMATE-ETIOLOGY, as described above in Section 3.2.2.1, also affects the formation of adequate, coherent hypotheses.

The final diagnosis is chosen from the collection of hypotheses formed using the rules of coherence and adequacy; often the basis for that decision is not just the scores of the individual components of each hypothesis, but the probability of their postulated interrelationships.

3.4.7 Symptom-Centered vs. Disease-Centered Processing

The mode of carrying out these four stages of processing changes gradually as a diagnosis proceeds. At the beginning of a diagnostic session, much of the emphasis is on the triggering phase, as the diagnostician is searching for some explanation for the findings and is willing to explore many possibilities. At the same time, the global assembling stage concentrates more on coherence and less on adequacy; often early in a diagnostic session, a doctor will have several coherent partial hypotheses, no one of which accounts for all the data. I call this symptom-centered processing, as each new finding is considered as a potential suggestion of new diagnoses. As a doctor invests more time and computation in a few hypotheses, however, inertia sets in - the triggering stage may be skipped altogether. The emphasis is instead on the inclusion of each new finding in new adequate hypotheses derived from those which existed

before the addition of that finding. The disposing stage assumes greater importance, as one of the basic activities in this mode of processing is attributing findings to already-considered hypotheses. Only rarely is a totally new hypothesis generated. This is disease-centered processing, because each complex hypothesis is considered in turn with regard to the new finding and modified to include it in some way or another. In the protocol, this type of processing was evident when the last few symptoms were added; HYPERTENSION CHRONIC was assumed to be caused by HYPERTENSION ESSENTIAL and added in to the other hypotheses, rather than other diseases being hypothesized to account for it.

3.4.8 Toward A Paradigm

The theory proposed here has been decidedly influenced by the fact that it is modeling medical knowledge and the diagnostic process. However, as discussed in Chapter 1, medical diagnosis can be thought of as just one example of a recognition problem, a category into which many other AI problems fall. What general points can we make at this point, before examining the details of the diagnostic algorithm, which might be applicable to a wider range of AI problems? The points below should provide the reader with a framework for reading the following chapters.

Most central to the theory described here is the idea of

several stages of processing, most particularly the distinction between a local stage and a global stage. The important idea is that we can't figure out how to account for all symptoms at once, so that smaller subsets of them have to be dealt with by more local hypotheses, which are then combined. In vision, for example, a telephone on a desk should be recognized by a two-step local-global procedure; first each component is recognized by itself, then the more global connection ON (which we might compare to COMPLICATED-BY) is used to combine the two in a complete "diagnosis."

Equally central is the notion of separating out a disposing phase, which attempts to account for symptoms as simply as possible and a triggering step which chooses the candidates for local evaluation. The notion of processing phases, so popular in compiler design, has been largely neglected in AI paradigms, often because the different stages were so interdependent. Perhaps, however, a more valuable approach is to start out with distinct processing stages, making the assumption that they don't interact - and then adding inter-stage communication as it becomes necessary. The dividing line between local and global may, as here, be somewhat arbitrary (cf. the distinction between findings and elementary hypotheses), but is useful as a first-order approximation.

An interesting side effect of considering different processing stages is the notion that the same piece of knowledge may be used in two different stages - and represented differently for use locally or

globally. Throughout this thesis, I refer to this as local compilation of global knowledge. Thus, the combination of hematuria and vaginal discharge may immediately trigger URINARY-TRACT-INFECTION COMPLICATED-BY VAGINITIS (in the triggering phase) or the COMPLICATION connection may not be discovered until the global assembling phase, where it is subsumed under more general methods. This multi-level representation of knowledge should be a valuable paradigm to follow.

Another important idea which clearly relates to other areas of AI research is that of a changing mode of processing, as described in the previous section. Traditional AI programs often make a distinction between top-down (disease-centered) and bottom-up (symptom-centered) approaches, but any one program operates in one mode (be it top-down, bottom-up or a combination) throughout its task. A gradual shift from one mode to another may be very relevant to language understanding. Mitch Marcus (personal communication) has considered a similar phenomenon in working on his parser. In the earlier stages of recognition, group structures (noun, verb etc.) are triggered by individual words or patterns; as the larger sentential structure is built, fewer new structures are triggered and more attention is paid to accounting for smaller details (e.g. agreement).

3.5 Summary

We have noted that the structure of the medical knowledge

necessary to do medical diagnosis is essentially a cause-effect net. The effects, called findings, have a structure which consists of a main-concept and a set of one or more property-values. When a real piece of data is asserted to the system, an attempt is made to fit it into various slots or finding specifications; several relationships between an actual finding and a finding-specification are possible: sufficient, further, insufficient and contradictory specification. The process of fitting is complicated by time-relationship considerations, including ABRUPT-ONSET-DISEASE and EPISODIC-DISEASE, as well as the obvious BEFORE, AFTER etc.

We also need relationships like CAUSE, COMPLICATION, and DEVELOPS-INTO between the causes, which are called elementary hypotheses. These relationships will play an important part in the global stage of processing. Elementary hypotheses may be related to more and less specific etiologies by CHOICE-SET and ISA links, respectively. Elementary hypotheses also have properties associated with them, such as EPISODIC-DISEASE and a TIME-INDEX, both of which help to correctly interpret RECURRENT-SYMPTOMS.

Finally, we quickly surveyed the processing necessary to come up with a final diagnosis. We first try to dispose of the newly-added finding by attributing it to an already-established etiology. Triggering is the next step, creating active instantiations of previous inactive hypotheses. Local evaluation determines which of the active hypotheses are to be accepted, which rejected and which

deferred. Global assembling tries to combine many of the local hypotheses into a more complex one which is both coherent because the ways hypotheses can be combined are limited and adequate to account for all the data.

With these preliminaries down pat, we should be able to look more closely at the various stages of the diagnostic process and determine how heuristics in each of them serve to keep a lid on the number of concurrently active hypotheses.

Chapter 4 - Local Evaluation

Before tackling larger questions of the coherence of hypotheses and the complications of their interactions, we need to have a method for evaluating an elementary hypothesis - usually a single disease or syndrome - in isolation. The actual mathematical method used and numerical scores generated are not of ultimate importance; using the scores comparatively to decide which hypotheses to continue actively pursuing and to guide the formation of larger, more complete hypotheses is clearly more crucial. However, the consideration of local evaluation brings up some conceptual issues relating to disease-centered vs. symptom-centered information, the role of each in a doctor's developing expertise, and how symptom-centered information is central to limiting active hypotheses. This chapter also expands the concept of slice to that of extended slice, which takes into account ISA links and age and sex specifications and explores a few examples of hypothesis-limiting information which this expansion brings up. The following discussion is concerned both with the structure of the medical knowledge and the processes which use that structure.

4.1 More on the Complete Theory - and How It Fails

4.1.1 Expectations vs. Evidence

The data we are faced with in designing a medical diagnosis system is a collection of signs, symptoms and properties of patients and a smaller collection of possible diagnoses. The task of any theory of medical diagnosis is to elucidate the correlations between these two types of entities so that, in a particular case, we may choose the most likely diagnosis. Some of the correlations are primitive, immediately distillable from data, while others are derived through more complex calculations. The conditional probability of a symptom given a disease is such a primitive correlation; I have called such numbers EXPECTATIONS. In Bayesian terms they are P(S/D) (read "the probability of S given D"), where S is the symptom and D the particular disease in question. These are the figures which are set forth, at least in words, in chapters in medical books on particular diseases. For example, these descriptions of symptoms of AGN:<Strauss and Welt 63>

> "Gross hematuria is one of the most common initial symptoms and occurs in more than one-third of the patients."
>
> "Edema is one of the most common presenting symptoms of the disease and is found in the great majority of patients."

"Hypertension occurs in the majority of cases."

"High fever and chills occur infrequently during the acute phase."

More expert doctors can validate these figures in their own experience or figure them out if they've forgotten them by thinking back over the last fifteen cases of a particular disease they saw and "counting" how many of them exhibited the symptom in question.

If P(S/D) = 1, then we know that <u>every</u> patient suffering from the disease D exhibits symptom S; we call this a NECESSARY EXPECTATION. The absence of S rules out D in this case, because P(-S/D) = 0, where -S means "not S." If P(S/D) = 0, in strict Bayesian terms, then no patient who has disease D exhibits symptom S and the appearance of S would rule out D. Our interpretation of such a correlation will be different, however - see section 4.2.2 below on relevant symptoms.

These correlations, however, are all <u>disease-centered</u>; that is, they spring directly from the description of a disease. Medical education is generally organized around such disease descriptions and thus a newly-graduated medical student can more easily describe a typical case of AGN or tell how gonorrhea is transmitted than name all the diseases in which hematuria might occur. However, more useful diagnostic information is <u>symptom-centered</u>, since a diagnosis proceeds from symptoms to diseases. It may be that the process of deriving more symptom-centered information characterizes much of a doctor's

movement toward expertise.

This other more sophisticated type of information is what I have termed EVIDENCE; in Bayesian terms, it is the conditional probability of a disease given a symptom, or P(D/S). The calculation of this EVIDENCE correlation between a disease and a symptom takes into consideration two dimensions besides the probability in the other direction (P(S/D));

1. the other diseases which can possibly account for the symptom in question and
2. the commonness of occurrence of each of the relevant diseases - their a priori probabilities.

For example, only some fraction of the people afflicted with glomerulitis have red blood cell casts, but there is no other disease which can cause this finding, so it is SUFFICIENT EVIDENCE for glomerulitis. Similarly, only some people with common colds get watery eyes, but colds are so common as to warrant suspecting one whenever a patient presents with watery eyes. The Bayesian formula for deriving these reverse probabilities (under certain assumptions which are discussed below) exhibits these two considerations. < Feller 68>

$$P(D_j|S) = \frac{P(D_j\ S)}{P(S)} = \frac{P(S|D_j)\ P(D_j)}{\sum_k P(S|D_k)\ P(D_k)}$$

$P(D_j/S)$ is the probability that symptom S is accounted for by disease D_j; $P(D_j S)$ is the probability that both D_j and S occur simultaneously, so it takes into account the commonness of D_j ($P(D_j)$), as well as the probability of S appearing in a patient with disease D_j . The denominator is the probability of S's occurring at all - a number which is derived by considering every other disease, its chance of accounting for the symptom and its a priori probability. The above equation is often referred to as "Bayes' rule for the probability of causes," so it is clearly at least conceptually applicable to our problem, which is one of cause and effect.

The complexity of this formula makes plausible the idea that part of a doctor's expertise lies in the translation of knowledge from the disease-centered mode to the symptom-centered mode. The compiled information about other diseases represented in one EVIDENCE assertion is considerable. What the EVIDENCE information really represents is the compilation of global information for use locally. The concept of local compilation of global knowledge is a crucial one in this thesis; it is exemplified most clearly both here and in the next chapter on non-linearity. Notice that this transformation is not a simple cross-index, since relative values must be attached to each of the pointers from symptom to disease; exceptionally high or low values on these pointers may be specially treated, as explained below.

4.1.2 Violations of the Assumptions of Bayesian Methodology

At first glance, it might seem that only EXPECTATIONS are necessary to do medical diagnosis. If the various symptoms are independent, if P (S_1 / S_2 D) = P (S_1/ D), where S_1 and S_2 are symptoms and D is a diagnosis, then we could just multiply the probabilities or their complements, according to whether the symptom occurred or not. By comparing all the products, we could choose the most likely diagnosis. This is precisely the "complete" theory referred to in Chapter 1. The formal assumptions underlying the Bayesian formula for deriving and using conditional probabilities as well as some common-sense practicalities point out why this is not feasible, as well as providing guidelines for more reasonable approaches.

1. The exclusive use of EXPECTATIONS would require evaluating the presence or absence of every symptom with respect to every possible hypothesis. In comparison with the concept of <u>perfect information</u> in game-playing, we may call this situation <u>perfect deduction</u>. Combinatorily and cognitively, this is clearly an impossible situation. We need a way to choose a smaller set of hypotheses to consider at any one time; multiplying and comparing products on that subset would then be more feasible. If the proliferation of hypotheses were the only problem, we might consider

the methods presented here an approximation to the Bayesian approach, suitably modified to fit in time and space limitations inherent in computers and humans. However, there are more basic problems with the probabilistic theory.

2. The derivation of the formula 4.1 relies on the various diseases' being exhaustive (i.e. the only possible causes of the symptoms in question) and mutually exclusive. If we choose a large enough selection of diseases, they may be an exhaustive set (we can guarantee this if we allow an UNKNOWN etiology), but we can certainly never hope to achieve a mutually exclusive set of causes, for this would necessitate all diagnoses including one and only one disease. In medicine, the most interesting and frequent conclusions include two or more diseases which may be related (as COMPLICATIONS, CAUSES etc.) or even unrelated.

3. The straight probabilistic approach gives us no straight-forward way to represent the temporal course of diseases or to take that data into account in the evaluation of the likelihood of a disease's occurring. The theory presented here proposes the beginning of a method for dealing with temporal material, although it has certainly not solved all the problem.

4. Any program which is going to be clinically useful and usable will have to explain the methods and data it used in reaching its conclusions to the physician using it. A response like "pulmonary embolism = .5, tuberculosis = .4" is not at all useful, since a doctor

must have more information about the reasons behind those numbers to feel comfortable about treating a patient.

5. Another assumption the derivation of formula 4.1 makes is that the symptoms' occurrence is independent. This is certainly not the case and several examples later will illustrate the types of interactions between symptoms which can occur on a local level. Probability theory does afford us a way to handle such non-linearities by finding separate values for P(S S /D) where they are relevant. Even though the probabilistic handling of the interaction situation is messy, we must realize that the interaction is in the data, not in the method, and thus any algorithm we devise will have to deal with the non-linearities. Therefore, although this is a place where the first-order probabilistic model breaks down, it is not a reason for rejecting probabilistic approaches.

The above reasons have led to the formulation of the following theory of local evaluation, reflecting the structure of an expert's knowledge and maintaining in particular the concepts of EVIDENCE and EXPECTATION developed above in the Bayesian framework.

4.2 The Expert's (Heuristic) Theory

4.2.1 What Should A Theory Do?

A brief digression is necessary here to discuss some general characteristics of hypothesis-evaluation. The consideration of any

hypothesis must take into account two relationships between the hypothesis and the data: I shall call these validity and sufficiency.

The validity of a disease hypothesis has to do with how many of the symptoms it often causes are present (EVIDENCE) and how many findings which are expected to be present are instead absent (VIOLATED EXPECTATIONS). I call the findings used in the local evaluation of a hypothesis relevant symptoms; this concept is expanded upon below. The degree of sufficiency of a hypothesis is determined by how many of the abnormal symptoms present it can account for, or be considered a cause of. Unaccounted-for symptoms lead to a search for more complex hypotheses which can account for all of the findings; these hypotheses may include more than one elementary hypothesis. This process is a more global one and is dicussed both below and in Chapter 6.

4.2.2 Relevant Symptoms

Recall from Chapter 3 above that I have called disease or syndrome nodes of the disease-symptom graph elementary hypotheses. The description of each disease (elementary hypothesis) mentions only a small number of the possible symptoms which might be encountered in a diagnostic session. These are the symptoms which can be accounted for by this diagnosis and whose presence or absence is thus relevant to its validity. As explained in chapter 3, we have called a disease and its group of relevant symptoms a slice. Symptoms not mentioned in

a disease's slice do not affect its validity score. Thus the STREP-INFECTION slice mentions ASLO-TITER HIGH, THROAT-CULTURE POSITIVE, PENICILLIN GIVEN, and FEVER, but contains nothing about HYPERTENSION or EDEMA. It's important to remember that not mentioning a symptom in a disease's slice doesn't mean that it never occurs concurrently with that disease, only that the disease can not be thought of as a cause for that symptom. HYPERTENSION can occur in a patient who also has a STREP-INFECTION if that patient is suffering from AGN. Part of the diagnostic problem is partitioning the symptoms into (not necessarily disjoint) subsets, each of which can be accounted for by an elementary hypothesis. Several of these can in turn be combined into a complete coherent hypothesis which accounts for all the symptoms.

4.2.3 A Scoring Algorithm

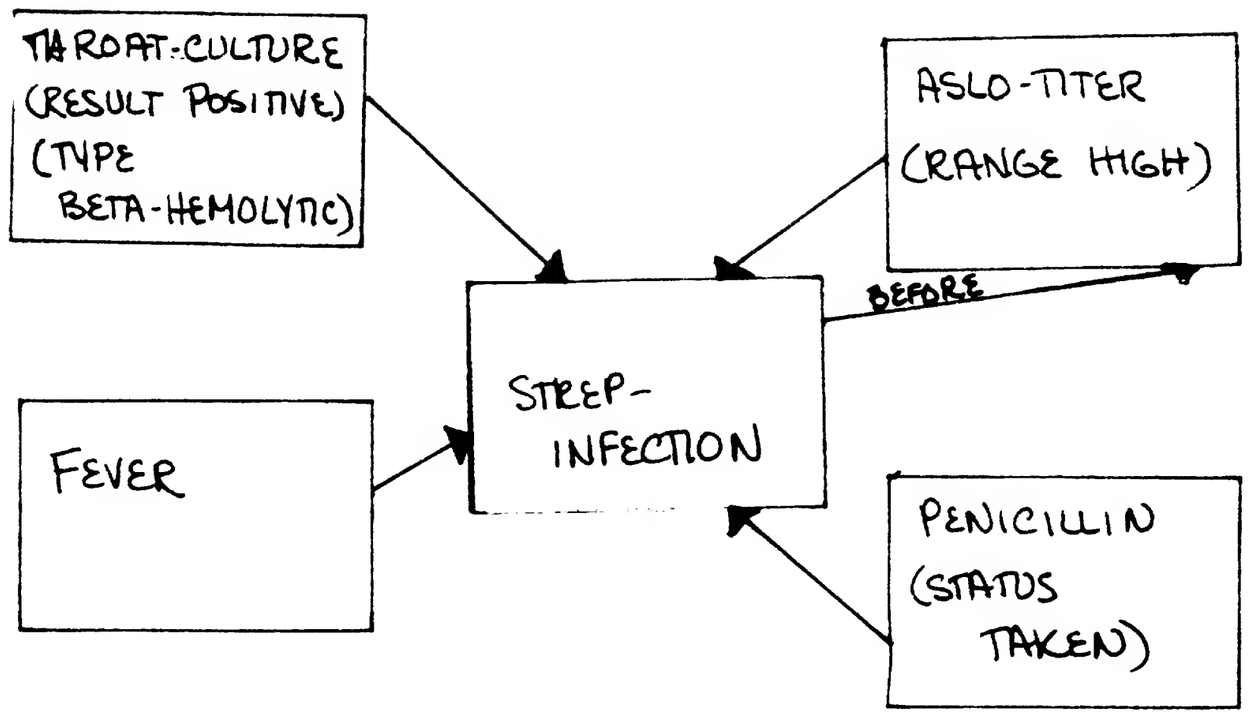
A local scoring algorithm must take into consideration both positive and negative contributions to the current hypothesis. In general the presence of relevant symptoms will add to the validity score of an elementary hypothesis, while their absence will subtract from it. The presence of FEVER will add to the validity of STREP-INFECTION, while its absence will subtract.

The theory we originally developed called for four levels of EVIDENCE (SUFFICIENT, STRONG, MODERATE, WEAK) and four of EXPECTATION

(NECESSARY, STRONG, MODERATE, WEAK). Because of the amount of medical expertise necessary to come up with any numbers at all and insufficient experimentation with any scoring system, the numbers in the examples below are derived mainly from Dr. Steve Pauker's estimates. The exact values of these numbers are rather unimportant; for now, consider STRONG, MODERATE and WEAK to be 1.0, .5, and .25, respectively, positive for EVIDENCE and negative for EXPECTATION. See Diagrams 4-1 and 4-1a for two examples of slices and their associated relevant symptoms; the diagrams themselves contain only the symptom and disease specifications, while the EVIDENCE and EXPECTATION strengths are listed separately.

The EXPECTATIONS used here do not always correspond exactly to the simple Bayesian probabilities introduced above (P (S/D)); rather, they may also be local compilations of global knowledge like the EVIDENCE strengths. For example, that PENICILLIN TAKEN is just a WEAK EXPECTATION in STREP-INFECTION does not imply that we don't expect that fact to be present, but merely that if it is not, our faith in the diagnosis is not shaken much. Clearly, some other criteria may come into play in the estimates of these values - perhaps the physician's intuition about the seriousness or importance of the symptom.

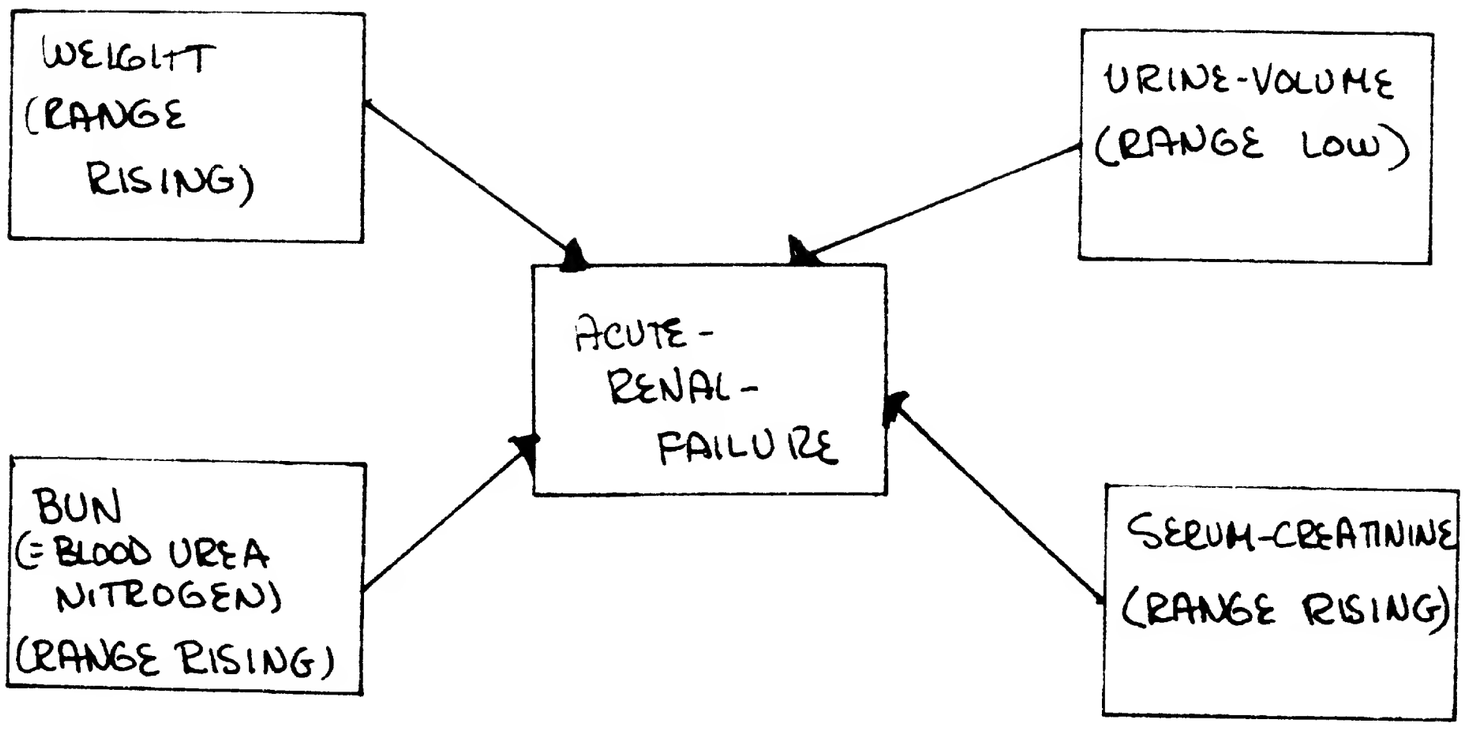
Both EVIDENCE and EXPECTATION strengths may be derived values, but it is clear that doctors also have the "raw" data available. Knowing how common a particular symptom is in a given disease really



RELEVANT SYMPTOM	EVIDENCE-STRENGTH	EXPECTATION-STRENGTH
THROAT-CULTURE (RESULT POSITIVE) (TYPE BETA-HEMOLYTIC)	SUFFICIENT	STRONG
FEVER	MODERATE	MODERATE
ASLO-TITER (RANGE HIGH)	SUFFICIENT	STRONG
PENICILLIN (STATUS TAKEN)	STRONG	WEAK

(BEFORE STREP-INFECTION (ASLO-TITER (RANGE HIGH)) (INTERVAL (WEEKS 1.5)))

Diagram 4-1: STREP-INFECTION SLICE, PASS 1



RELEVANT SYMPTOM	EVIDENCE-STRENGTH	EXPECTATION-STRENGTH
URINE-VOLUME (RANGE LOW)	STRONG	MODERATE
SERUM-CREATININE (RANGE RISING)	SUFFICIENT	NECESSARY
BUN (RANGE RISING)	STRONG	NECESSARY
WEIGHT (RANGE RISING)	MODERATE	WEAK

Diagram 4-1a: ACUTE-RENAL-FAILURE SLICE, PASS 1

is their primary knowledge and is the form most often used for explanation and certainly for communicating with other doctors. In addition, the verifiable facts must be available as the basis for debugging, a process which we hope physicians go through often. Some more comments on the uses of this disease-centered information are included in Chapter 7.

Using only four distinct strengths lumps together a possible infinitude of values into a few larger categories. I have allowed only a small number of EXPECTATION and EVIDENCE strengths to limit the amount of numerical complexity in scoring and because of some general arguments, which will not be detailed here, aginst the use of the full range of real numbers between 0 and 1 as possible values for correlations between entities. In another paper, I have argued that on scales such as TALLNESS, AGE etc., there are probably a handful of discrete categories (VERY-TALL, PRETTY-TALL etc.) into which measurements fall. <Rubin 73> For more exact comparisons, there are most likely dual comparisons like (TALLER-THAN HARRY JOHN), but there is no guarantee that such orderings form a complete ordering or are even consistent. A similar situation probably exists in medicine. I have insisted on limiting the different strengths of EVIDENCE and EXPECTATION pointers to a handful. In addition, there are specific assertions of the form (MORE-LIKELY DISEASEX DISEASEY) (usually "given a few symptoms") which differentiate between diseases which have symptoms in common in cases where our limited numerical scoring may be

too coarse to draw the line. These assertions are used by the global assembling phase. We would not expect, of course, a complete ordering to arise from these specialized assertions. Only certain ones will ever be relevant and as a doctor's expertise develops, the useful comparisons will be generated and remembered.

There are two obvious methods for evaluating an elementary hypothesis. In either case, first add up the EVIDENCE for the hypothesis and subtract the violated EXPECTATIONS to obtain the raw score. In order to normalize it, this raw score can either be divided by the highest total possible score the hypothesis could have or by the highest total score it could have taking into account just the symptoms mentioned. For example, suppose a patient had BUN (RANGE RISING) but URINE-VOLUME (RANGE NORMAL). The raw score for ACUTE-RENAL-FAILURE would be 1 - .5 = .5. Dividing by the total score would yield .5/ 1 + 1 + .5 + 1 (for these purposes SUFFICIENT and STRONG EVIDENCE count the same) = .5/ 3.5 = 1/ 7. Call this the total-related score. Dividing by the highest score achievable with just information on OLIGURIA and BUN yields .5/ 1 + 1 = 1/ 4. Call this the included-related score. Because symptoms are discovered serially, we can never assume that more information about a particular symptom won't be forthcoming. Thus, the second scoring algorithm seems to take into account the fact that information is incomplete, while the first compares what we know with a situation in which information on more symptoms is available. The included-related score

seems more appropriate for the early stages of a diagnosis, while the total-related score, because it assumes that information is somewhat complete, might come into play later. In real-life situations, doctors try not to be faced with incomplete-information situations by asking questions to determine the status of relevant symptoms, but of course some information may not be obtainable. For the limited part which numerical scores play in this theory we will use the total-related score to accept a hypothesis, if it ever becomes 1. We will use the included-related score to reject a hypothesis, if it ever becomes less than 1/8 (this threshold is experimentally untested and I don't stand by it). Of course, hypotheses can also be accepted by the introduction of SUFFICIENT EVIDENCE or rejected by the violation of a NECESSARY EXPECTATION.

The scores the system finally comes up with range over the rationals; this seems contradictory to my original argument that only a few values of EVIDENCE and EXPECTATION are desirable. In fact, people probably have a very imprecise system for combining probabilities and the scores they end up with are certainly not as exact as .875. Figuring out how such a human system might work is still a major research topic.

4.2.4 Scales of Property Values

Not all symptoms can be only present or absent; there are

RELEVANT SYMPTOM	EVIDENCE-STRENGTH	EXPECTATION-STRENGTH
URINE-VOLUME (RANGE LOW)	STRONG	MODERATE
SERUM-CREATININE (RANGE RISING) (RANGE HIGH)	SUFFICIENT MODERATE	NECESSARY
BUN (RANGE RISING) (RANGE HIGH)	STRONG WEAK	NECESSARY
WEIGHT (RANGE RISING)	MODERATE	WEAK

Diagram 4-2: ACUTE-RENAL-FAILURE SLICE, PASS 2

degrees of severity for many symptoms which complicate the preceding analysis of EVIDENCE and EXPECTATIONS. SERUM-CREATININE may not be RISING, but it may be HIGH; this finding does not contribute as much EVIDENCE, but neither does it detract from the hypothesis of ACUTE-RENAL-FAILURE as SERUM-CREATININE NORMAL would. Similarly, HEMATURIA can be GROSS or MICROSCOPIC and these different severities may contribute differently to a given elementary hypothesis.

A solution is to allow different properties to have different EVIDENCE strengths and to assume EXPECTATIONS come into play with the introduction of a symptom which is contradictorily specified (see Chapter 3) with respect to the symptom-specification in the slice. Thus, in most cases where differing severities are included in the EVIDENCE column, the EXPECTATION amount would be subtracted if the given symptom were absent or the test result normal. See Diagram 4-2 for the ACUTE-RENAL-FAILURE slice redone to take this into account.

I have followed the convention here of only mentioning abnormal findings explicitly in a slice, since the original definition of relevant symptom was a symptom which could be accounted for by the disease in question. This approach is "global" in the following sense: Consider the status of blood pressure in focal glomerulonephritis (FGN), which is usually normal. We have two possible ways to represent this :

1. Consider HYPERTENSION ABSENT as EVIDENCE for FGN and its absence (= the presence of hypertension) a violated EXPECTATION of the FGN

hypothesis

2. Leave HYPERTENSION out of the list of FGN's relevant symptoms altogether and hope that a hypothesis which accounts for its presence would differentially "win out" if it were to appear.

I have chosen the latter approach, since we clearly cannot mention in every disease's slice all the symptoms it can't account for. The explicit mention of one or two of them is an example of compiled information (see Chapter 5 and section 4.3.3.1) which add to the effectiveness of the system, since it requires a global view to know which non-relevant symptoms it is important to include in a disease's slice. The inclusion of non-relevant symptoms in a slice provides a mechanism for explicitly rejecting a hypothesis, rather than allowing it to remain active and only be rejected later by comparison with other hypotheses which account for more symptoms.

4.2.5 Extended Slices

The slices I have been considering have included all those symptoms which are relevant to the given disease; they are represented in the diagrams as those symptoms connected by one pointer (actually an abbreviation for a pair, EVIDENCE and EXPECTATION) to the corresponding disease. However, there are other factors to be considered in the local evaluation of an elementary hypothesis.

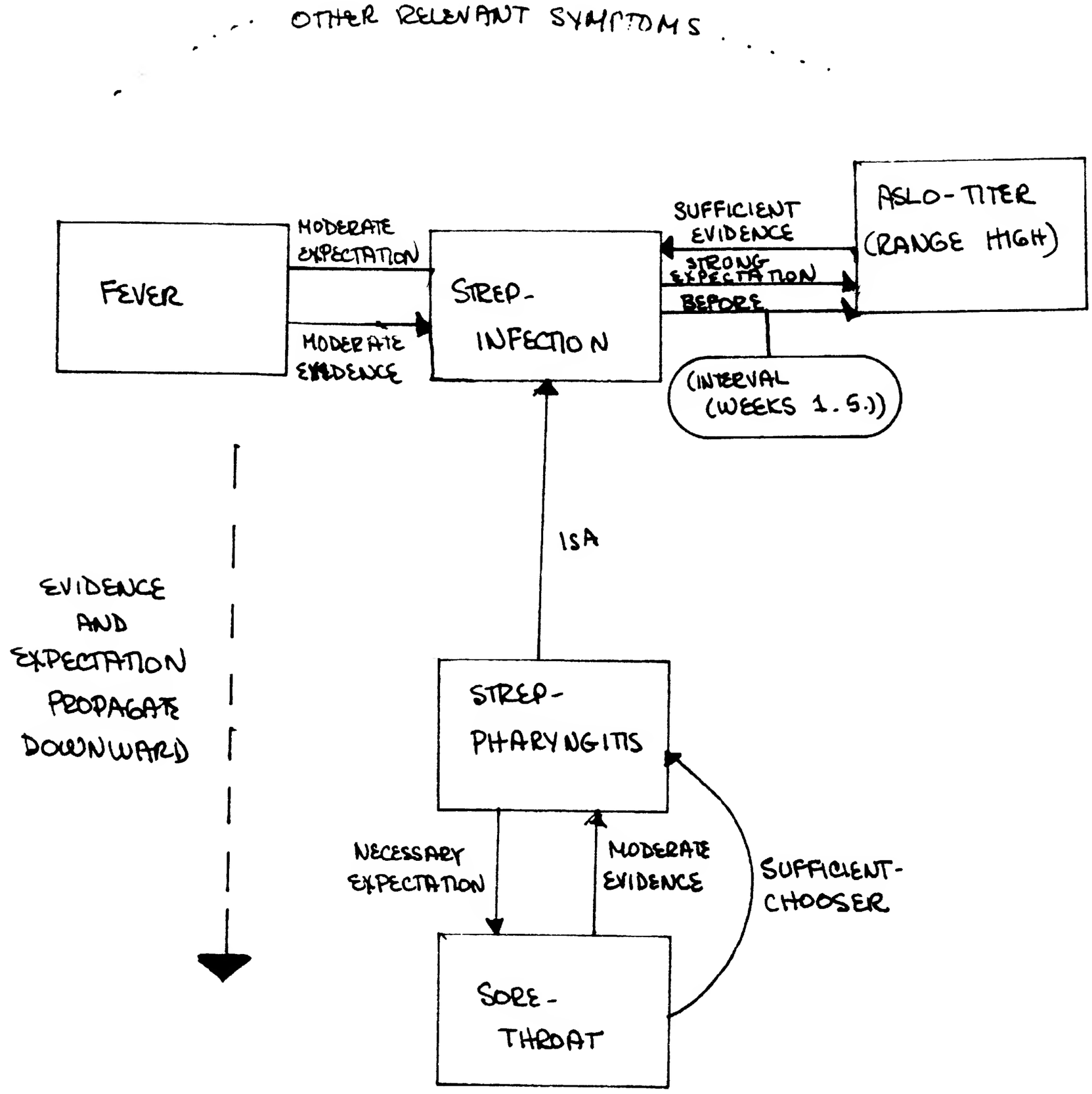


Diagram 4-3: THE MECHANICS OF ISA-LINKS

4.2.5.1 ISA links

ISA is a way to express hierarchies. DISEASE-X ISA DISEASE-Y means that DISEASE-Y is a general classification and DISEASE-X is an example of that class. If X ISA Y, then if the diagnosis of X is confirmed, the diagnosis of Y is confirmed. For example, SCARLET-FEVER ISA STREP-INFECTION, so if we are satisfied that the patient has SCARLET-FEVER, either by assertion or by an appropriate test, then we are also satisfied that he or she has a STREP-INFECTION. Similarly, FGN, LGN and AGN are all examples of GLOMERULITIS. In these cases, the more general hypothesis (e.g. STREP-INFECTION) is called the category and the more specific ones the examples (e.g. STREP-PHARYNGITIS). The complete set of examples of a more general disease is called the CHOICE-SET of that disease; the CHOICE-SET of STREP-INFECTION is (STREP-PHARYNGITIS STREP-SKIN-INFECTION SCARLET-FEVER). A CHOICE-SET is intended to be mutually exclusive; it may also be exhaustive - if so, it is so marked as in Diagram 4-4. This provides the additional information that if the category is an accepted hypothesis and if all but one of the examples are rejected, the remaining one may be accepted.

Those symptoms which are relevant only to the example are attached by EVIDENCE and EXPECTATION relationships to just that disease, while symptoms which are more generally relevant to the

category are related to the higher classification. For example, STREP-PHARYNGITIS ISA STREP-INFECTION. A SORE-THROAT is MODERATE EVIDENCE for STREP-PHARYNGITIS, as well as a NECESSARY EXPECTATION. FEVER, however, is related to all types of STREP-INFECTION as MODERATE EVIDENCE and MODERATE EXPECTATION, so it appears in the more general slice. (See Diagram 4-3) Clearly, these EVIDENCE and EXPECTATION relationships propagate unchanged "down" the ISA link to STREP-PHARYNGITIS, so that we can regard the hierarchical structure like a shorthand which eliminates the need for re-representing these relationships in each slice corresponding to a disease which ISA STREP-INFECTION.

Sometimes a symptom is evidence for one and only one of the members of a CHOICE-SET; thus, if the category is already definitely diagnosed, the presence of that symptom is sufficient to CHOOSE one of the examples. For example, LVH and RETINOPATHY HYPERTENSIVE are both SUFFICIENT CHOOSERS for HYPERTENSION CHRONIC in the category HYPERTENSION; the complete CHOICE-SET is HYPERTENSION CHRONIC and HYPERTENSION ACUTE. Since a symptom may appear in more than one CHOICE-SET, or perhaps, via hierarchy, a whole series of them, the CHOOSER relation is defined relative to a particular CHOICE-SET. Clearly, this information is always derivable by looking at the other members of the CHOICE-SET and their associated relevant symptoms. In this case, only one of the examples will be able to account for the symptom. However, in general, determining that may sometimes involve

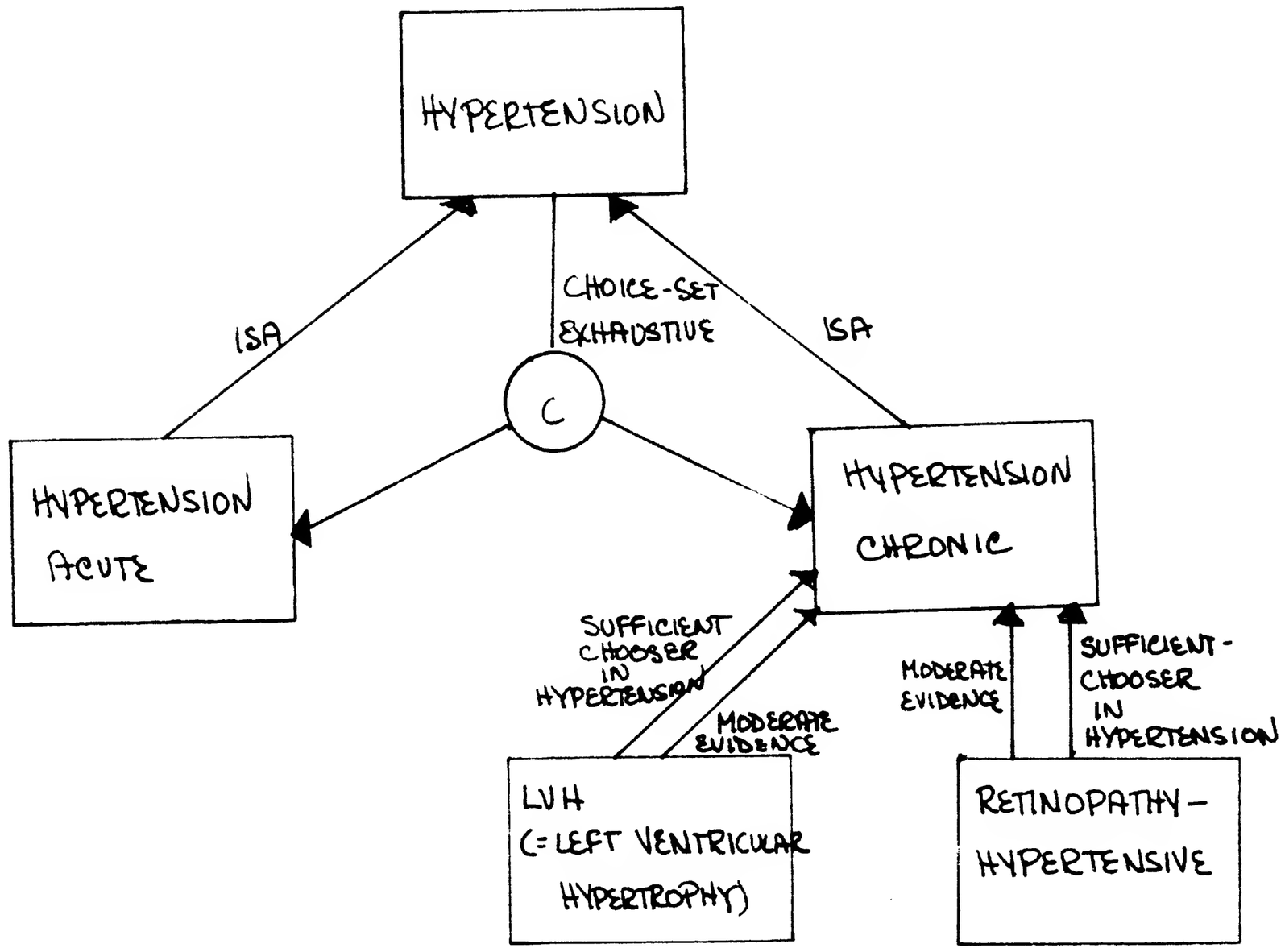


Diagram 4-4: HYPERTENSION CHOICE-SET AND CHOOSERS

a lot of computation - nor can we disregard the possibility that some other etiology might account for the symptom. So the inclusion of a SUFFICIENT CHOOSER pointer really represents another local representation of global knowledge, only in this case "global" does not mean all possible diseases, but rather a small subset. These are most helpful in places where we can expect the category to be confirmed first and a choice between its examples to be made only afterward; this <u>is</u> the case with the category HYPERTENSION, and the relevant structure is shown in Diagram 4-4. This type of local compilation is clearly a powerful mechanism; designating a particular symptom a SUFFICIENT CHOOSER in an artificially-constructed CHOICE-SET of often-confused diseases is a possible method for differential diagnosis. Future investigators should be on the look-out for such structures. The role of ISA links in the global evaluation mechanism in discussed in Chapter 6 in the overall discussion of the coherence of hypotheses.

4.2.5.2 Age and Sex

The age and sex of a patient obviously play a large part in the diagnosis of illness. Certain diseases, such as measles and mumps, are predominantly childhood diseases; cystine stones occur mainly in children, while uric acid stones occur mainly in adult males. Therefore we must allow age and sex to act as evidence for or

against various hypotheses. This information represents local compilation of global knowledge about a priori probabilities, corresponding to the local compilations of symptom-disease correlations mentioned above. Representing the various effects of age and sex differences on hypotheses necessitates changing slightly our view of properties with various values. Recall from above that various properties on a concept may change the amount of EVIDENCE it contributes, but we have made the EXPECTATION a single amount which is subtracted when the symptom is contradictorily specified. However, we may want to assert that different values of AGE (INFANT, CHILD, ADOLESCENT, etc.) may subtract different amounts from a hypothesis. To do this we want to be able to associate with each value on a property scale a single strength and the designation of positive or negative. We may have in an extended slice of MEASLES:

AGE

INFANT +MODERATE

CHILD +STRONG

ADULT -MODERATE

or AGN

AGE

INFANT -WEAK

CHILD +STRONG

ADULT -WEAK

This is not really inconsistent with the notation used above, as we could translate every EXPECTATION into an explicit mention of the symptom with the negation of a property, ABSENT or NORMAL attached and an associated negative value. I have chosen to retain the notion of EXPECTATION for its linguistic value - a NECESSARY EXPECTATION or a VIOLATED EXPECTATION are easy to conceptualize - as well as to make clear the connection between slices and a pure Bayesian framework. To show the isomorphism, however, I have included Diagram 4-5, which illustrates the STREP-INFECTION and ACUTE-RENAL-FAILURE slices in the notation which corresponds to the designation of AGE above. For a more complete discussion of contradictory specification and related issues, see Chapter 3.

4.3 Cutting Down on Active Hypotheses

The major thrust of the theory I have been developing is to explain and exemplify heuristics by which the number of hypotheses actively being considered at any particular time can be minimized. What contributions do the elements of the extended slices mentioned above and the concepts of EVIDENCE and EXPECTATION make to this overall goal?

RELEVANT SYMPTOM	EVIDENCE/EXPECTATION STRENGTH
THROAT-CULTURE	
(RESULT POSITIVE) (TYPE BETA-HEMOLYTIC)	+SUFFICIENT
CONTRADICTORY	-STRONG
FEVER	
(PRESENCE PRESENT)	+STRONG
(PRESENCE ABSENT)	-WEAK
ASLO-TITER	
(RANGE HIGH)	+SUFFICIENT
(RANGE NORMAL)	-STRONG
PENICILLIN	
(STATUS TAKEN)	+STRONG
CONTRADICTORY	-WEAK

RELEVANT SYMPTOM	EVIDENCE/EXPECTATION STRENGTH
URINE-VOLUME	
(RANGE LOW)	+STRONG
(RANGE NORMAL)	-MODERATE
SERUM-CREATININE	
(RANGE RISING)	+SUFFICIENT
(RANGE HIGH)	+MODERATE
(RANGE NORMAL)	-NECESSARY
BUN	
(RANGE RISING)	+STRONG
(RANGE HIGH)	+WEAK
(RANGE NORMAL)	-NECESSARY
WEIGHT	
(RANGE RISING)	+MODERATE
(RANGE NORMAL)	-WEAK

Diagram 4-5: STREP-INFECTION AND ACUTE-RENAL-FAILURE SUCES, PASS 3.

4.3.1 Search, Plausible Move Generators and Triggers

The "complete" theory of medical diagnosis as described above is analogous to an exahustive search; each hypothesis is examined in turn, with little motivation for choosing one hypothesis before or instead of another. To reduce otherwise intransigent search spaces, as in chess, the concept of Plausible Move Generation has been introduced in Artificial Intelligence. (one example of its use in chess is in <Greenblatt 69>). A Plausible Move Generator specifies just those moves which are worthwhile pursuing, leaving out the vast majority of possible moves. Similarly we need a mechanism which suggests only a few elementary hypotheses to be considered at one time. Obviously, the same factors taken into consideration in determining EVIDENCE from EXPECTATIONS are crucial here - the other hypotheses which could account for the symptom, their a priori probabilities and the probability of the symptom occurring in each disease.

As explained in more detail at the end of Chapter 3, a hypothesis may be <u>activated</u> by one of its <u>triggers</u> (this terminology comes originally from Minsky, Winograd and other people who are working on <u>frames</u> - see Chapter 7.) A hypothesis which is not active is not being currently considered or evaluated. Triggers for an elementary hypothesis are generally a subset of those symptoms which are STRONG or SUFFICIENT EVIDENCE. In the STREP-INFECTION slice,

THROAT-CULTURE POSITIVE BETA-HEMOLYTIC and ASLO-TITER HIGH are triggers; in the ACUTE-RENAL-FAILURE slice, OLIGURIA, SERUM-CREATININE RISING and BUN RISING are all triggers. FEVER, by itself, on the other hand, probably doesn't trigger anything because it can be accounted for by so many diseases. This selective activation of hypotheses is one way to control the number of diseases being actively considered at any time. Notice that this use of triggers is certainly a heuristic device, since the diagnosis for the particular case on hand may not be one of those triggered.

In the protocol in Chapter 2, one of the most striking features is the activation of the Polycystic Kidney Disease hypothesis by the mention of familial nephritis; even though three other hypotheses were being considered and none of them was in serious trouble, the force of the suggestion of familial nephritis was sufficient to make the doctor seriously entertain that hypothesis.

4.3.2 Negative Activation

Another way to keep the number of active hypotheses low is to get rid of unlikely ones. The protocol also contains an example of negative activation - the consideration and immediate rejection of an elementary hypothesis. In this case, the knowledge that PROTEINURIA LIGHT was a symptom was sufficient knowledge to reject NEPHROTIC-SYNDROME, since PROTEINURIA HEAVY is a NECESSARY EXPECTATION

for the diagnosis. From a computational point of view this seems a wasted effort, since N-S (NEPHROTIC-SYNDROME) wasn't being actively considered anyway; it wasn't activated by the previous mention of HEMATURIA. If, later, a trigger for N-S had been added, the hypothesis could have immediately been evaluated and rejected. However, we are dealing here with a cognitive system with limited memory. There is a chance that later on in the diagnostic session, the doctor will have forgotten some of the specific symptoms, but will be able to remember that N-S has been rejected. He or she is remembering the results of a deduction, rather than the facts used to determine it. This is especially important because N-S is a commonly-occurring malfunction.

4.3.3 Making Definite Decisions

It's always nice to be able to make a definite decision! Being able to accept an elementary hypothesis or completely reject it lessens the cognitive load of a particular diagnostic situation. The presence of SUFFICIENT EVIDENCE allows the doctor to confirm a hypothesis; here again is an example of how the translation of disease-centered information into symptom-centered information increases the efficiency and effectiveness of a diagnostic process. The violation of a NECESSARY EXPECTATION (unless an EXCUSE is available - see Chapter 5) allows the diagnostician to reject an

elementary hypothesis. Also, we have the convention that if an elementary hypothesis' total-related score reaches 1, we consider it confirmed and when its included-related score reaches 1/8, we can consider it rejected. These, too, are somewhat heuristic methods, since any case may be atypical. So far, these are the only ways we have of making a definite decision about a hypothesis; below we extend this set heuristically in order to make the process more efficient.

4.3.3.1 Unaccounted-for Symptoms

Recall that the findings in a disease's slice are those which it can account for. A symptom which is present but cannot be accounted for by a candidate hypothesis is a phenomenon which is handled on a more global level (see Chapter 6); sometimes that symptom will cause the hypothesis to be rejected, sometimes it will result in a more complex hypothesis, much of which depends on the stage of the diagnosis, comparative validities and sufficiencies of hypotheses and other global characteristics of the situation. One type of compiled heuristic information is the inclusion of non-relevant symptoms in a disease's slice so that their presence can reject the hypothesis immediately, without recourse to global methods and comparing hypotheses. For example, the presence of RED-BLOOD-CELL-CASTS rules out the diagnosis of SICKLE-CELL-TRAIT. What rejecting symptoms is it important to include in a slice? Clearly, these are a form of

differential diagnosis; the symptoms which are most necessary to include are those which distinguigh the disease in question from other diseases with which it shares many symptoms. Rejecting symptoms are, like EVIDENCE pointers themselves, a compilation of global evidence for local use: the RED-BLOOD-CELL-CAST example above really contains the fact that there is no coherent hypothesis (see Chapter 6) containing SICKLE-CELL-TRAIT which accounts for the casts; that information is really global, involving several different elementary hypotheses and connections between them - but it has been condensed into a single assertion which can be used locally.

Another example of explicit rejection of an elementary hypothesis is the interaction of HEMATURIA and PROTEINURIA in GLOMERULITIS. Both HEMATURIA and PROTEINURIA are relevant symptoms for GLOMERULITIS - both can be accounted for. However, the conjunction of specific severities of each of them explicitly eliminates the GLOMERULITIS hypothesis. The combination of HEMATURIA GROSS and PROTEINURIA LIGHT cannot be accounted for by GLOMERULITIS, so it is rejected. This example is noteworthy because it is another case of definite rejection of a hypothesis, as well as the first example we have come across of interaction (in this case, between two symptoms), the major topic of Chapter 5.

4.3.3.2 A Priori Probabilities

Different diseases, of course, have different probabilities of occurrence, called a priori probabilities. The age and sex of a patient affect this probability profoundly. Combining age, sex and disease leads to a useful number representing the probability of the disease occurring in a patient of particular age and sex. If this number is especially low, we may consider it 0 for heuristic purposes and put the hypothesis on the DEFERRED-LIST. The protocol in chapter 2 contains a clear example of this phenomenon: the presenting symptom was HEMATURIA, which is a trigger for G-U-TUMOR, among other renal diseases. However, the probability of a G-U-TUMOR in a 31-year-old woman is so small that the doctor did not actively consider the hypothesis at that time. Clearly, if more and more symptoms suggestive of TUMOR were to arise, the hypothesis would have to be resurrected, but at this point the age-sex-a-priori probability is so low that the hypothesis is rejected. We call this phenomenon premature rejection, as the diagnostician refuses to consider a hypothesis even though a symptom triggers it and even though he or she doesn't have a mathematically correct reason to reject it. Certainly this is a heuristic measure and prone to error, as some women of 31 have G-U-TUMORS; it is another method of limiting the number of active hypotheses by only considering those which seem promising.

It is obviously worthwhile for a physician to compile such

probabilities in his or her movement toward expertise, as the age and sex of a patient are always available. Dr. J. P. Kassirer has often said that age and sex are two of the most important facts in a diagnosis; given age, sex and presenting symptom he can often predict the final diagnosis. Recently, he was so disturbed at the diagnosis of pulmonary embolus in a 30-year-old woman that he ordered a re-evaluation of a lung scan which had been interpreted in support of a pulmonary embolus. The sex-age-a-priori probability of the diagnosis in the patient was so low as to cast doubt on even the most "reliable" evidence; the pathologists interpreted the lung scan as normal the second time around, vindicating Dr. Kassirer's intuitions.

4.5 Summary

This chapter, besides explaining the details of a scoring algorithm for elementary hypotheses, has also described some elements of the theory which aid in limiting the number of hypotheses actively considered at one time. In addition, a distinction between local and global information is beginning to emerge.

Our original prototype for local knowledge was the EXPECTATION of symptom given disease; this merely involves a single symptom and a single disease. In contrast, global information is that which requires knowing facts about more than one disease (often <u>all</u> diseases being considered). Many of the examples of hypothesis-limiting

mechanisms have required the "compilation" of global information into a form which is locally-usable, within the context of the evaluation of a single elementary hypothesis.

The transformation of medical knowledge from a disease-centered mode to a symptom-centered mode was seen as a large step in a doctor's developing expertise, as well as a prototypical example of the compilation of global knowledge for local use. Computing the strength of the EVIDENCE pointer for FEVER in STREP-INFECTION, for example, from EXPECTATIONS requires knowledge about all the other diseases which could cause FEVER - clearly global knowledge; the final result - that a FEVER is MODERATE EVIDENCE for a STREP-INFECTION - is usable locally, in the evaluation of the STREP-INFECTION elementary hypothesis.

Designating a subset of these EVIDENCE pointers as triggers afforded us a way to assure that only hypotheses which were actively suggested by a present symptom would be active. This contrasts with the complete theory, in which a hypothesis is active unless it is "ruled out" by the presence or absence of some symptom.

Although slices have been defined to mention only relevant symptoms, or those which can be accounted for by the disease, another mechanism we have considered for cutting down the number of active hypotheses is the explicit inclusion of symptoms which are not relevant in a slice in order to reject that hypothesis. Without such explicit direction, the doctor (or a program) might search for a CAUSE

or COMPLICATION of the elementary hypothesis which might account for the symptom. For example, RED-BLOOD-CELL-CASTS rule out SICKLE-CELL-TRAIT; there is no use searching for a more complex hypothesis containing SICKLE-CELL-TRAIT which might account for the casts.

A similar localization of knowledge occurred in the context of ISA links, where a symptom could be designated a CHOOSER of a particular example of a category; the example here was LVH and RETINOPATHY HYPERTENSIVE, which allowed the definite choice of HYPERTENSION CHRONIC. Making a choice allows the elimination of the other members of a category's CHOICE-SET, again reducing the number of active hypotheses.

A final mechanism noted was that of premature rejection; the placing of a triggered elementary hypothesis on the DEFFERED-LIST because of its a priori probability given the age and sex of the patient. Whle this does not really represent a local expression of global information, it is certainly a heuristic measure which allows fewer simultaneously active hypotheses. The primary example here was the dismissal of TUMOR in the case of a 31-year-old woman.

Chapter 5 continues the investigation of heuristic methods for making the task of diagnosis possible by cataloging symptom-symptom interactions which serve as such mechanisms.

Chapter 5 - Non-Linearities

The theory presented so far has been, with the exception of a few hints along the way, a linear one. Such a theory assumes that subparts of a problem can be treated separately and independently and the solutions to those subproblems combined without alteration. One manifestation of linearity in medical diagnosis is the assumption that the strengths of EVIDENCE and EXPECTATION can be evaluated independently for each symptom relevant to a disease. Another manifestation is the clustering of symptoms into pathophysiological states with the assumption that they may be evaluated equivalently regardless of what disease-context they appear in. This chapter will give examples of interactions which contradict the linear theory, as well as indicating how some of those interactions may be compiled in an expert's diagnostic strategy for greater efficiency.

5.1 Non-Linearity: Recent Investigations

The concept of non-linearity has recently been recognized in Artificial Intelligence as a circumstance which pervades many problem domains and problem-solving approaches. Two recent theses on debugging have identified overlooked interactions as a source of program bugs.

Sussman's research <Sussman 73a> on programs which build block structures noted that successfully accomplishing a conjunctive goal like

(AND (ON B C) (ON A B))

where A, B and C are blocks, requires being aware of an interaction between the subparts. Accomplishing (ON A B) and then trying to do (ON B C) creates a problem because B, which has A on top if it, can't be picked up to be put on C until A is removed. Sussman's program HACKER solves the problem by reordering the steps to eliminate the interference. The bug arises from the assumption that plans for solving the subgoals can be combined without any provision for interaction between them.

Goldstein's thesis <Goldstein 74> studies the domain of Turtle programs which draw simple geometric pictures. In trying to discover the plan of a program - the way it relates to the model of the picture it is supposed to draw - Goldstein first tries a linear plan. Such a plan assumes that the parts of the picture as defined by the model are drawn in succession, often in some geometrical progression like top to bottom. Even a linearly planned program has interactions which the programmer must notice; these are the interfaces between the subparts in which the direction and heading of the Turtle must be changed in preparation for the next step. If a linear plan does not reflect the relationship between program and model, Goldstein tries non-linear plans, such as an insert plan, which interrupts the code for one part

of the model with that for another part.

The examples of non-linearity described below correspond, to a large extent, to complicating the disease/finding network described above in a particular way: by making the entities in the finding nodes complicated expressions involving findings, rather than the individual findings themselves. A linear theory assumes, for example, that if symptom A contributes STRONG EVIDENCE for some disease and symptom B is MODERATE EVIDENCE, the amount of evidence contributed by the concurrent presence of both A and B is some linear combination of STRONG and MODERATE - most obviously, their sum, suitably normalized. Assigning a different value to (AND symptomA symptomB) indicates that there is some correlation between the symptoms which affects the amount of evidence their conjunction represents. For a down-to-earth example, consider: the fact that I have mud on my left shoe is evidence of my having taken a walk in the woods; so is having mud on my right shoe. Having mud on both shoes, however, is not twice as much evidence as having it on one, as the two findings are highly correlated; if one occurs, the other one does too. In this case, the evidential contribution of (AND A B) would be less than the sum of A and B. Logical combinations may also include operators like OR and NOT.

Some of the specific medical examples listed below are taken from Steve Pauker's study of EDEMA and related complaints. I have attempted to catalogue the types of interactions I have noted both in

his material and in mine. The first category, which affects the local evaluation stage, and the second, which affects the global assembling stage, are declarative interactions; the information represented there is necessary for diagnosis to happen at all. The third and fourth, however, which occur during the triggering and local evaluation stages respectively, are heuristic interactions. They relate more to the process of diagnosis and represent heuristics for keeping the number of active hypotheses at a reasonable level.

5.2 Declarative Interactions in Local Evaluation

5.2.1 HEMATURIA and PROTEINURIA in GLOMERULITIS and G-U-TRACT-BLEEDING

Both HEMATURIA and PROTEINURIA are EVIDENCE for GLOMERULITIS and G-U-TRACT-BLEEDING. However, their relative severities differ in these two hypotheses. In G-U-TRACT-BLEEDING, we expect the ratio of HEMATURIA to PROTEINURIA to be near that in whole blood; for HEMATUIRA GROSS we expect PROTEINURIA LIGHT (100-1000 mgs. in 24 hours). In GLOMERULITIS, on the other hand, there should be relatively more PROTEINURIA than in G-U-TRACT-BLEEDING; for PROTEINURIA MODERATE (1000mgs. - 4 gm. 24-hour urine protein), we would expect HEMATURIA MICROSCOPIC or LIGHT, but most likely not GROSS. The approach I have taken to this interaction is to specify for each disease or state which combinations would rule it out. Thus (AND (HEMATURIA GROSS) (PROTEINURIA LIGHT)) precludes GLOMERULITIS, while (AND (HEMATURIA

LIGHT) (PROTEINURIA HEAVY)) precludes G-U-TRACT-BLEEDING. Clearly, this does not represent all of the knowledge a diagnostician has about this comparison. He or she also knows facts like, "If the HEMATURIA/PROTEINURIA ratio is lower than in whole blood, it is more likely GLOMERULITIS than G-U-TRACT-BLEEDING, but we can't rule either one out." Some kind of gradient exists; at one of its endpoints are the combinations ruling out GLOMERULITIS and at the other are those combinations which rule out G-U-TRACT-BLEEDING. In between are various states in which the relative likelihoods change. This information seems to be centered around the entity HEMATURIA/PROTEINURIA-RATIO and may be used by comparing it with the ratio of red blood cells to protein in whole blood. Thus, the information used in processing may be represented as differential diagnoses like

```
(MORE-LIKELY G-U-TRACT-BLEEDING GLOMERULITIS
        (WHEN (HIGH HEMATURIA/PROTEINURIA-RATIO)))
```

Such assertions might be used by the global assembling phase which is the only one which has access to many elementary hypotheses at once. This is a representational problem, however, which needs to be investigated further.

5.2.2 Sex-related characteristics: TESTICULAR ATROPHY

Clearly, certain findings are sex-related. TESTICULAR ATROPHY, a finding in CIRRHOSIS (a progressive destruction of viable liver tissue), is one such symptom which obviously only occurs in males. The absence of this finding shold not detract from the hypothesis of CIRRHOSIS in a woman, so the local evaluation should take place as if the finding weren't relevant to the hypothesis at all. This interaction, although it affects local evaluation, is certainly specific only to the symptom, not to any of the elementary hypotheses to which it is relevant. It is therefore stated globally only once as (SEX-RELATED TESTICULAR-ATROPHY MALE) and the local evaluation mechanism checks for such exceptions before actually evaluating each hypothesis. Although the fact is stated globally, its use does not imply a global search for caveats on each hypothesis, as the SEX-RELATED assertions are indexed under the symptom name and are thus immediately retrievable.

The above example is really an interaction between a FACT and a SYMPTOM; the parallel relationship between a FACT (sex of the patent) and an elementary hypothesis is often evidenced in the age-sex-a priori probability of the disease. For example, PREGNANCY would never be considered as a possible cause for nausea or fatigue in a man. In this respect, sex and race are similar; SICKLE-CELL-TRAIT is not a possible etiology for HEMATURIA (or anything else) in a white

person. As explained above, these a priori probabilities can cause hypotheses to be rejected immediately, as is appropriate.

5.2.3 BLOOD-UREA-NITROGEN (BUN) and SERUM-CREATININE in ACUTE-RENAL-FAILURE

Both BUN and SERUM-CREATININE levels are indicators of renal function, as the kidneys filter the materials both tests measure and remove them from the body. If the levels are normal, the kidneys are functioning well; if either one is elevated, it indicates renal failure. These two measures occur together in renal disease, so the interaction between the two findings when they are both present resembles the muddy shoes example from above; a situation in which both levels are elevated is not too much more evidence for ACUTE-RENAL-FAILURE than SERUM-CREATININE elevated, if the BUN is unknown. However, if the BUN is elevated, but the SERUM-CREATININE isn't, ACUTE-RENAL-FAILURE is precluded. In fact, another diagnosis is suggested - a necrotizing tumor - in much the same way that mud on only one shoe might suggest my having hopped through the mud. Such DIFFERENTIAL-DIAGNOSIS pointers will be discussed in more detail below.

5.2.4 Excuses: PENICILLIN and ASLO-TITER in STREP-INFECTION

The ASLO (anti-streptolysin-O) titer often rises several weeks after a person has had a STREP-INFECTION, indicating that the body is fighting the infection with antibodies. Taking PENICILLIN to combat the infection, however, often squelches the antibody response. If a doctor or diagnostic program were actively considering STREP-INFECTION, ASLO-TITER (RESULT NORMAL) would represent a violated expectation. An excuse is sometimes available for the absence of an expected finding; in this case PENICILLIN (STATUS TAKEN) would excuse a normal ASLO-TITER (as well as contributing some evidence of its own to the hypothesis of STREP-INFECTION). The STREP-INFECTION hypothesis is evaluated as if ASLO-TITER were not a relevant symptom when penicillin has been taken.

Another example of this type of interaction was alluded to in the protocol; RED-BLOOD-CELL-CASTS are expected in GLOMERULITIS. HEMATURIA GROSS can explain their reported absence, however, because lots of red blood cells in the urine can obscure the casts when they are looked for under a microscope.

Sometimes the excuse is not a FACT like PENICILLIN GIVEN or a symptom like HEMATURIA GROSS, but a disease itself whose presence or absence must be determined by more complicated evaluation. For example, HYPERTENSION is NECESSARY for the diagnosis of HYPERTENSION CHRONIC, but a MYOCARDIAL INFARCTION (heart attack) can cause the

absence of HYPERTENSION in a chronically hypertense patient. If HYPERTENSION CHRONIC were being considered, but it was discovered the patient did not have HYPERTENSION, a coherent hypothesis would be formed which included MYOCARDIAL-INFARCTION as an excuse for HYPERTENSION ABSENT.

Sometimes, of course, the best strategy is to reject a hypothesis rather than to try to find an excuse for a perceived discrepancy. The decision whether to keep searching or to give up on a hypothesis is not an easy one. McDermott <McDermott 74> has developed a formalism for assimilating new and possibly contradictory information in a language-understanding system. His methods, which involve building a "ring" of related assertions which support or explain one another, include provisions for EXCUSES and other similar structures.

5.2.5 OR-clauses: CHEST PAIN in MYOCARDIAL-INFARCTION

Often several further specifications (see Chapter 3) of the same finding are evidence for the same disease and are basically mutually exclusive. Since only one of them will occur in any instantiation of the hypothesis, we should not consider the total possible score for the hypothesis to reflect the concurrent presence of <u>all</u> of those symptoms, but rather of one. For example, CHEST PAIN further specified as SQUEEZING, PRESSING, DULL or VERY-SEVERE is

STRONG EVIDENCE for a MYOCARDIAL-INFARCTION. In any specific patient, probably only one of these descriptors will apply. The effect desired can be obtained, of course, by constructing an OR clause containing the various further specifications, as

(PAIN (LOCATION CHEST)

(CHARACTER (OR SQUEEZING PRESSING DULL VERY-SEVERE)))

The interpretation of such a structure, which is crucial in explaining the course of a diagnosis, is that any ot the disjuncts can fill the slot and that one filler is all that is really expected.

Often differing severities call for a structure using OR.

SERUM-CREATININE (SEVERITY (OR HIGH RISING))

is evidence for ACUTE-RENAL-FAILURE, as

WEIGHT (RANGE (OR HIGH RISING))

is evidence for SODIUM-RETENTION.

5.2.6 Discontinuities in Evaluation: EDEMA and PROTEINURIA in NEPHROTIC-SYNDROME

The evaluation procedure outlined in Chapter 4 has a major discontinuity; a hypothesis can be accepted either by having its total-related score equal 1 or by the presence of some finding which is SUFFICIENT EVIDENCE for the disease. Often the conjoined presence of two or more findings is sufficient to confirm a hypothesis, even though any one of them alone would not be. This is the case with

EDEMA (SEVERITY MASSIVE) and PROTEINURIA (SEVERITY VERY-HEAVY) in NEPHROTIC-SYNDROME; the concurrent presence of both findings confirms this diagnosis. Symptoms which cannot be accounted for by the elementary hypothesis may be included in such an interaction as well, although they are not usually mentioned in the slice of a disease which could not cause them. (i.e. they are not relevant) For example, the presence of HYPERTENSION along with the absence of RETINOPATHY HYPERTENSIVE is sufficient to confirm HYPERTENSION ACUTE.

This sixth interaction type really moves out of the domain of predominantly declarative interactions to heuristic ones. Just as EVIDENCE pointers are a local compilation of global knowledge, so are these patterns which confirm hypotheses, for they implicitly include the information that no other disease can account for this particular collection of findings - clearly global knowledge. Their existence reflects again the importance of the doctor's being able to make definite decisions, to accept or reject an elementary hypothesis rather than just changing its score.

5.3 Context-symptom interactions

The clustering of symptoms into pathophysiological states which can then be evidence for many different diseases creates another type of interaction problem. Consider, for example, SODIUM-RETENTION. EDEMA (fluid retention in the tissues) of various sorts is evidence

for SODIUM-RETENTION, which in turn is caused by many different diseases such as acute glomerulonephritis (AGN) and CIRRHOSIS. The particular manifestation of EDEMA in CIRRHOSIS however is usually EDEMA (LOCATION GUT) (the medical term is ASCITES), while that in AGN is most often EDEMA (LOCATION FACE)

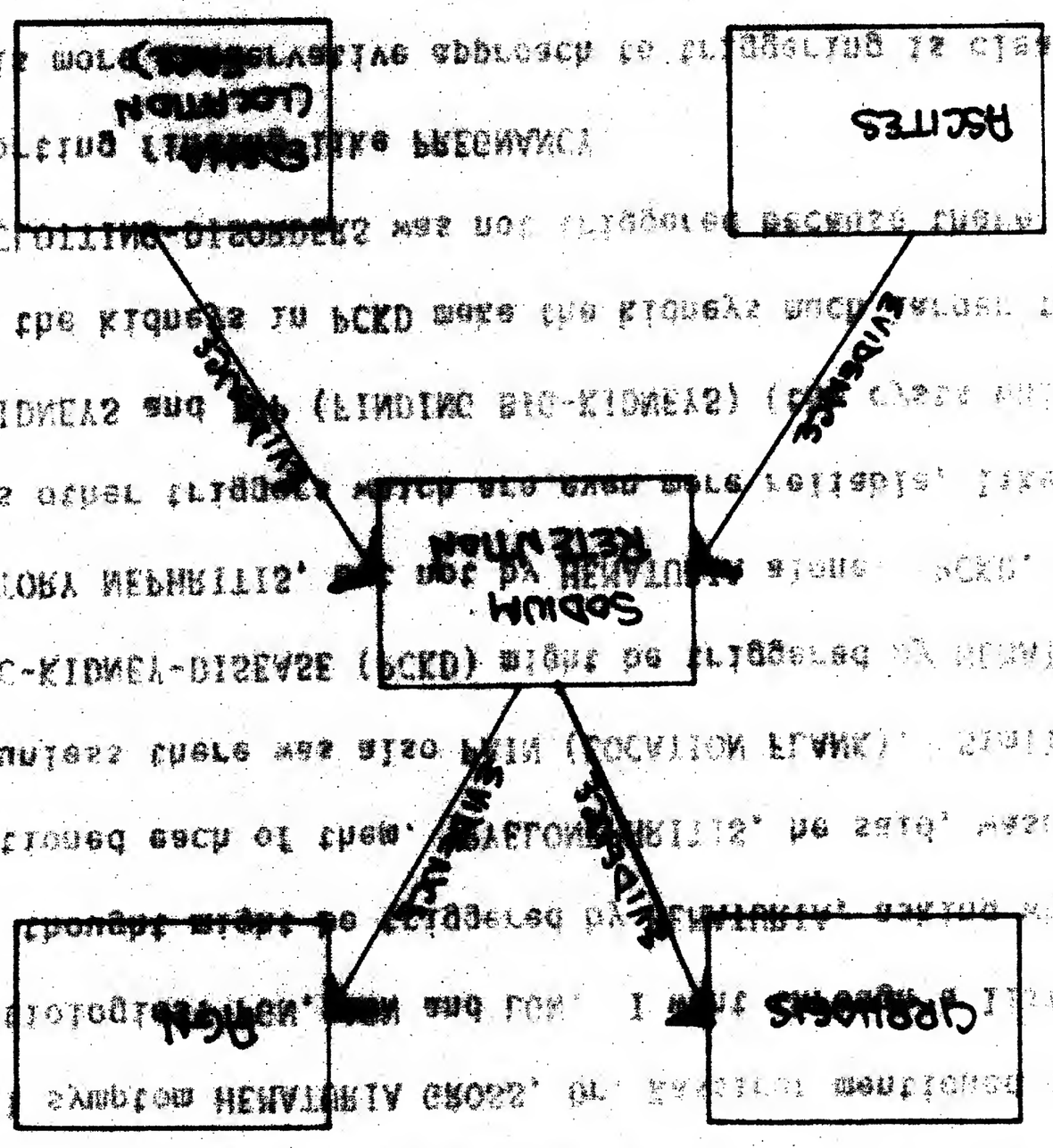
This situation has been called (by Sussman and myself) the "X" phenomenon, because in its most simple form it represents the relationship between two symptoms and two etiologies where a common cluster is interposed in the middle, as illustrated in Diagram 5-1.

In order to preserve the possibility of evaluating SODIUM-RETENTION independently (which is important, as it is often hypothesized by doctors as an intermediate step before triggering a specific disease), it is necessary to treat this situation specially in the global assembling phase. When a hypothesis is "put together" containing ASCITES, SODIUM-RETENTION and AGN, the assertion (OVERRIDE ASCITES AGN RARE) is noted, and the global hypothesis is deemed less coherent as a result. (Coherence is discussed in detail in Chapter 6.) This interaction represents information about the diseases and findings which is necessary for diagnosis, rather than heuristic information like the following types of interactions.

5.4 Heuristic Interactions in the Triggering Phase

The concept of trigger as introduced in Chapter 4 was that a

Diagram 5-1 Title "X"



single finding would activate certain hypotheses which commonly cause it. In going over the protocol in Chapter 2 with Dr. Kassirer, it became clear that this mechanism is too simplistic. When faced with the initial symptom HEMATURIA GROSS, Dr. Kassirer mentioned only three possible etiologies: FGN, AGN and LGN. I went through a list of diseases I thought might be triggered by HEMATURIA, asking why he hadn't mentioned each of them. PYELONEPHRITIS, he said, wasn't activated unless there was also PAIN (LOCATION FLANK). Similarly, POLY-CYSTIC-KIDNEY-DISEASE (PCKD) might be triggered by HEMATURIA and FAMILY-HISTORY NEPHRITIS, but not by HEMATURIA alone. PCKD, of course, has other triggers which are even more reliable, like PALPABLE-KIDNEYS and IVP (FINDING BIG-KIDNEYS) (the cysts which develop in the kidneys in PCKD make the kidneys much larger than normal.) CLOTTING-DISORDERS was not triggered because there was no other supporting finding like PREGNANCY.

This more conservative approach to triggering is clearly a heuristic aimed at controlling the number of active hypotheses. An alternative approach would be to allow HEMATURIA by itself to trigger many more elementary hypotheses and have certain ones become more and more likely as more supporting symptoms were added. Notice, also, that these instances of triggers using more than one symptom were discovered by the doctor's explaining his actions, rather than being directly derivable from his actions alone. Possible differences between explanations which come from more declarative information and

actions which may be compiled are discussed in more detail below.

The issue of multiple triggers is an important one and deserves more attention. It is possible that more than two symptoms may comprise a multiple trigger or that more complex logical combinations involving OR and NOT may be used. One or more of the conjuncts in a multiple trigger may be elementary hypotheses rather than findings. Although I haven't yet discovered examples of all these interaction types, it seems clear that they are possible and should be investigated further.

5.5 Differential Diagnoses

Two or more diseases may resemble each other in many of their crucial aspects; it is particularly important to be able to tell them apart. Medical textbook descriptions of diseases often contain a section labelled "differential diagnosis" which points out those findings which can differentiate between the diseases. This is probably one of the few places in such textbooks where "symptom-centered information" creeps in.

Besides being a possible pitfall in causing misdiagnoses, findings which are shared among diseases can also be used heuristically to avoid activating an undue number of hypotheses. Suppose diseases A and B share findings X, Y and Z, but are differentiated by Q's occurrence in A but not B. If X and Y are

present, we can consider B but not A, provided there is also a piece of heuristic information which activates A and rejects B if Q is discovered. This is the case with Acute and Chronic Glomerulonephritis (AGN and CGN), which share the symptoms of GLOMERULITIS (HEMATURIA, PROTEINURIA and RED-BLOOD-CELL-CASTS), as well as EDEMA (LOCATION FACE). CGN, however, expects HYPERTENSION CHRONIC and RENAL-FAILURE CHRONIC, while AGN exhibits HYPERTENSION ACUTE and RENAL-FAILURE ACUTE. Since AGN is more common, a doctor may consider it first and then switch to CGN if HYPERTENSION CHRONIC is discovered. Of course, the newly-suggested hypothesis must be evaluated itself, as some of the symptoms relevant to the first disease may not occur in the second at all. Suppose a doctor suspected AGN in a patient because of HEMATURIA, PROTEINURIA and a case of STREP-PHARYNGITIS three weeks earlier. The introduction of HYPERTENSION CHRONIC may cause CGN to be activated and evaluated, but CGN can't explain the STREP-INFECTION'S connection. Thus, the diagnostician is left with two possibilities: hypothesize CGN and consider the STREP-INFECTION unrelated or hypothesize AGN and consider the HYPERTENSION unrelated. On the basis of the findings presented here, there is no clear choice; more questions would have to be asked of the patient.

The notion of having a particular finding move consideration from one hypothesis directly to another has been examined recently in a paper on cube-recognition <Kuipers 74>. His system starts by

considering that the line drawing it is examining in fact represents a cube; the discovery of an angle which is too small sends it off to the "wedge" hypothesis which is then explored in greater depth.

5.6 Interpretation vs. Compilation

A theme which has run through the above discussion is that of local compilation of global knowledge. Because the concept of compilation appears in various forms, I will summarize some of the relevant examples and ideas here.

Global knowledge in this domain refers to information which requires knowing about more than one elementary hypothesis or disease. The first example of local compilation I noted was the EVIDENCE pointers (Chapter 4) which theoretically encompassed knowledge about all possible diseases which could cause a symptom; triggers were introduced as a subset of EVIDENCE pointers whose selection evidence a similar compilation of global knowledge. The multiple triggers suggested above for diseases like PYELONEPHRITIS and POLY-CYSTIC-KIDNEY-DISEASE are a further extension of the EVIDENCE idea; hypotheses are only activated if they are better candidates than many others and the multiple trigger idea allows the system to be even more selective than one trigger would permit. In fact, it is likely that the procedure that the doctor follows during an actual diagnostic session is even more compiled in the following sense: from age, sex

and presenting symptom, he or she is able to jump directly to a few possible diagnoses in a manner similar to access from a hash table. Therefore, given (SEX FEMALE) (AGE 31) and HEMATURIA (SEVERITY GROSS), Dr. Kassirer immediately responded with FGN, LGN and AGN. Later, however, he had to be able to explain to me why other choices were inappropriate, although that information was not contained in the compiled portion of the code. (See Section 2.1 for a discussion of the doctor's explanation mode.) For example, he did not actually activate RENAL-INFARCTION and then reject it because of a priori probability; rather, it was not in the hash table access list, so was not even considered. The declarative information is necessary as an explanation, if for nothing else. The idea of code existing simultaneously in many states along the declarative/procedural continuum has also been discussed by Winograd in the context of considering a design for a programming assistant <Winograd 74>.

The "hash table" of age, sex and presenting symptom certainly represents a local compilation of global knowledge, as its construction requires information about other (less likely) diseases. The discovery that expert doctors often jump to conclusions which, in fact, may turn out to be wrong is the major emphasis of an earlier paper by Sussman on medical diagnosis. <Sussman 73b>.

In the protocol there is also mention of the question of "level of evaluation" - that is, if HEMATURIA GROSS is a trigger for GLOMERULITIS, is just the GLOMERULITIS hypothesis activated, or are

all of its examples - AGN, LGN, etc - triggered as well? Here again there seems to be no uniform solution; Dr. Kassirer's "hash table" in this instance pointed specifically to three disease, rather than to the more general category GLOMERULITIS. In other cases, however, it makes more sense to activate only the category - G-U-TUMOR is a more sensible hypothesis than BLADDER-TUMOR, KIDNEY-TUMOR etc. if there is no information to differentiate between them. The exemplary hypotheses should be activated if some finding is added which would differentiate between them, such as an IVP showing a mass in the bladder. This "information-theoretic" approach often is not followed by doctors, who tend to jump to a more specific conclusion than is warranted, then make up for their undue haste by the use of differential diagnoses, as explained in Section 5.5. Note, however, that the presence of the same information in several forms makes the system (human or computer) less sensitive to the strategy selected (jumping to conclusions vs. entertaining more general hypotheses), because it has several procedural paths to any diagnosis.

One of the differential diagnoses cited above provided another, slightly different, example of local compilation of more global knowledge. First of all, every differential diagnosis mentioned is an example of such compilation, for each is based on the global knowledge that two diseases are similar in many respects, but different in at least one crucial way. But there is something else going on in the AGN-CGN case. Recall that seeing HYPERTNESION CHRONIC

when AGN is being considered makes CGN a good candidate. In addition, seeing a finding like RETINOPATHY HYPERTENSIVE which is STRONG EVIDENCE for HYPERTENSION CHRONIC should result in the CGN hypothesis being activated. Some techniques for making this more global connection are suggested in the next chapter on global assembling, but those techniques can be compiled as well. In AGN's slice there may actually be the differential diagnosis: RETINOPATHY HYPERTENSIVE => consider CGN. This is, of course, more efficient and direct as it represents a compilation of the chain of EVIDENCE pointers which connect RETINOPATHY HYPERTENSIVE to CGN (RETINOPATHY HYPERTENSIVE is EVIDENCE for HYPERTENSION CHRONIC which is EVIDENCE for CGN.) However, the intermediate information regarding HYPERTENSION CHRONIC must also be available for explanatory purposes. Again, we see special shortcuts being taken in compiled code, while the more declarative interpretable information must remain for explanation and possible debugging.

5.7 Summary

This chapter has examined the non-linear aspects of the theory, identifying both declarative interactions which are facts about medicine necessary for any diagnostic procedure and heuristic interactions which represent compiled shortcuts for performing diagnosis more efficiently and keeping the number of active hypotheses

to a minimum. Several different declarative interactions which have only local consequences were noted, as well as a declarative interaction dubbed the "X" phenomenon which had to be dealt with by the global phase of processing. Interactions in the triggering phase and in differential diagnosis were viewed as heuristic interactions, and some more general theory of interpretation and compilation was developed to provide a framework for these heuristics.

"The best explanation for any phenomenon is always the simplest one available that accommodates all the facts."

---The Exorcist

Chapter 6 - Global Assembly

Most of the diagnoses at which doctors finally arrive are not represented by a single elementary hypothesis. Patients often have more than one related or even totally unrelated diseases. A final diagnosis may be NEPHROTIC-SYNDROME COMPLICATION-OF GLOMERULITIS or, as we saw in the protocol, FGN and HYPERTENSION ESSENTIAL. Clearly we need some way to discover and specify these more complex hypotheses. In addition, we must be able to combine pathological states which are themselves elementary hypotheses into a larger hypothesis which postulates a single disease as a cause for all of them. These concerns are handled by the global assembly stage of processing. This chapter explores the various functions of a global phase of processing, a phase which has access to all of the nodes and links of the knowledge net, instead of just those which cluster around a single elementary hypothesis, as well as to global assertions which give information about more than one hypothesis. Like preceding chapters, this chapter also identifies some heuristics used in global assembly which help limit the number of concurrently active hypotheses. In

fact, most of the activities of this phase move toward reducing the number of hypotheses the physician must remember by unifying a group of them into a larger structure. This effort is obviously parallel to the use of elementary hypotheses themselves to organize data. In fact, many of the structures created by this step will be seen to have clear analogs in the local evaluation/elementary hypothesis sphere.

Most of the processes in this chapter are described in terms of matching a pattern (a template) and performing some action on the basis of that match. Clearly, the triggering and differential diagnosis actions described in the previous chapters could be similarly conceptualized in terms of pattern-matching and associated actions. There is definitely a unified theory lurking here; at this point, however, it is probably better to concentrate on explaining local and global processes separately and worry about unifying them later.

6.1 To Be or Not To Be...Coherent

Recall the arguments above in Chapter 4 regarding definite decisions: it's better to be able to definitely accept or reject a hypothesis rather than keeping track of its changing score or relative ranking among the other considered possibilities. The theory of coherence I have developed contains that assumption explicitly; a complex hypothesis is either coherent or not. At this point, there is

no coherence score associated with a complex hypothesis - only the symptom- and time-scores of the elementary hypotheses which are its components. A complete system would need some kind of back-tracking mechanism which would allow it to choose an "incoherent" hypothesis after discovering that no coherent hypothesis fit sufficiently with the data to be considered the final diagnosis. Further sections and examples will make this necessity clear.

6.2 Local Coherence of Time-Instantiations

The simplest kind of combination of elementary hypotheses actually occurs in the local evaluation stage and results in what I have called locally coherent hypotheses. Each separate occurrence over time of a symptom is interpreted as being caused by the same disease in evaluating the elementary hypothesis corresponding to that disease. In the protocol, for example, all occurrences of HEMATURIA were interpreted as symptoms of GLOMERULITIS in evaluating the GLOMERULITIS hypothesis; they were all interpreted as evidence of G-U-TRACT-BLEEDING in evaluating that elementary hypothesis. This amounts to joining all the time-instantiations of a particular elementary hypothesis into one composite hypothesis whose score is taken to be the average of the scores of all the time-instantiations. This clustering is schematically illustrated in Diagram 6-1. When I use the term "elementary hypothesis" below in this chapter, it may

refer to such a locally coherent hypothesis, for once the time-instantiations are combined, they may be thought of as one entity.

This heuristic may sometimes cause a doctor or diagnositc system to miss a diagnosis, since sometimes the same symptom is caused by different diseases on different occasions. In a recent simulated patient/doctor interaction (not the same one which was reported in the protocol), Dr. Kassirer was misled because one previous occurrence of the patient's hematuria was caused by a kidney stone, while all the others were symptoms of FGN. Dr. Kassirer did eventually discover that two separate and unrelated diseases were involved, but this was not his initial guess. We can see that a complete diagnostic system would have to be able to consider less coherent possibilities if the locally coherent hypotheses didn't work out. In the case cited, the presence of pain with one occurrence of hematuria, but not with the others was probably the crucial clue. Situations which require incoherent complex hypotheses such as these are commonly called "red herrings" - and may produce great anger and irritation when they are included in a clinical pathological conference (CPC) to trick the physician trying to diagnose the case.

6.3 Evidence Chains

Many of the mechanisms cited below are necessitated by the

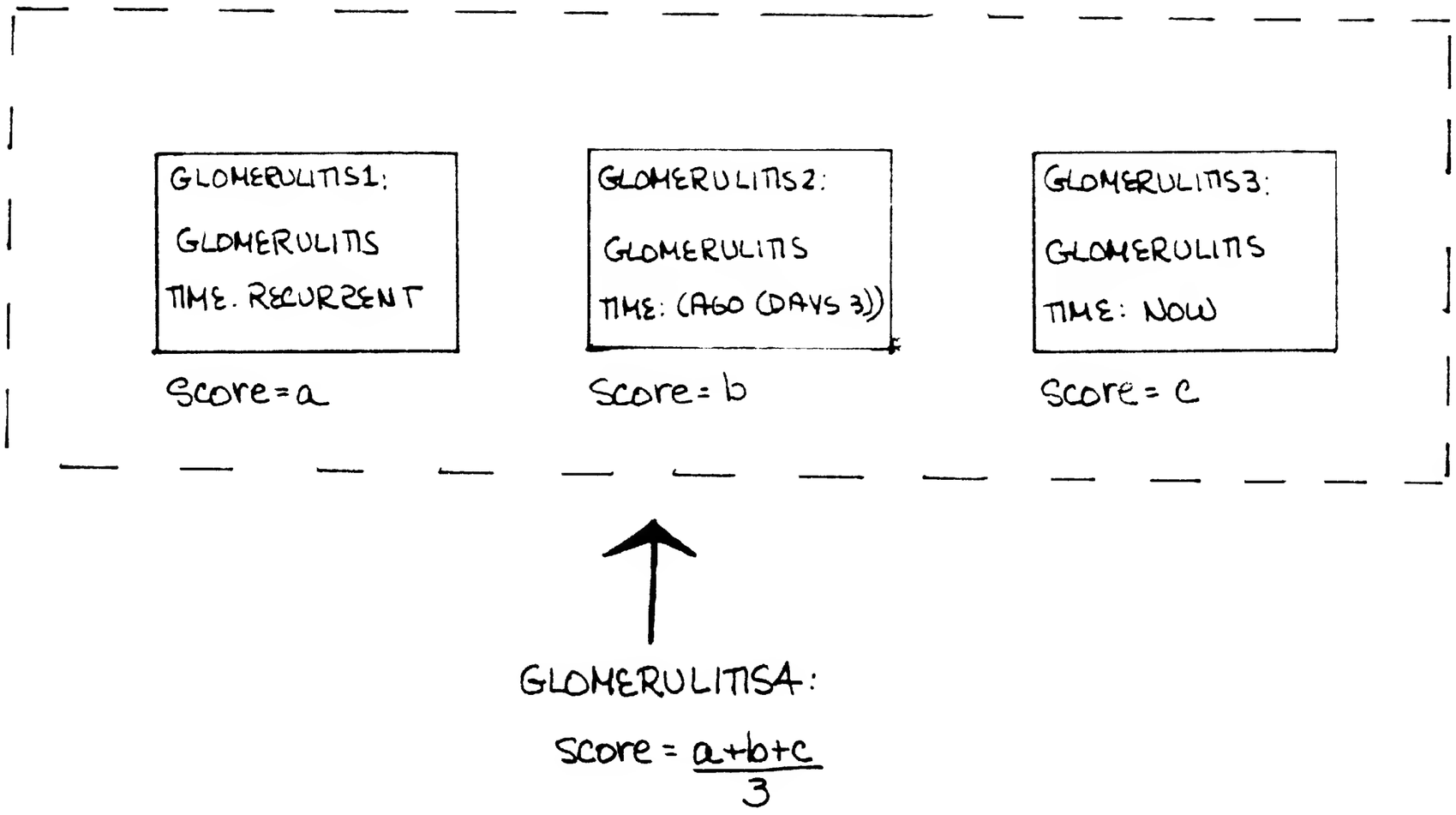


Diagram 6-1: Locally Coherent Hypothesis

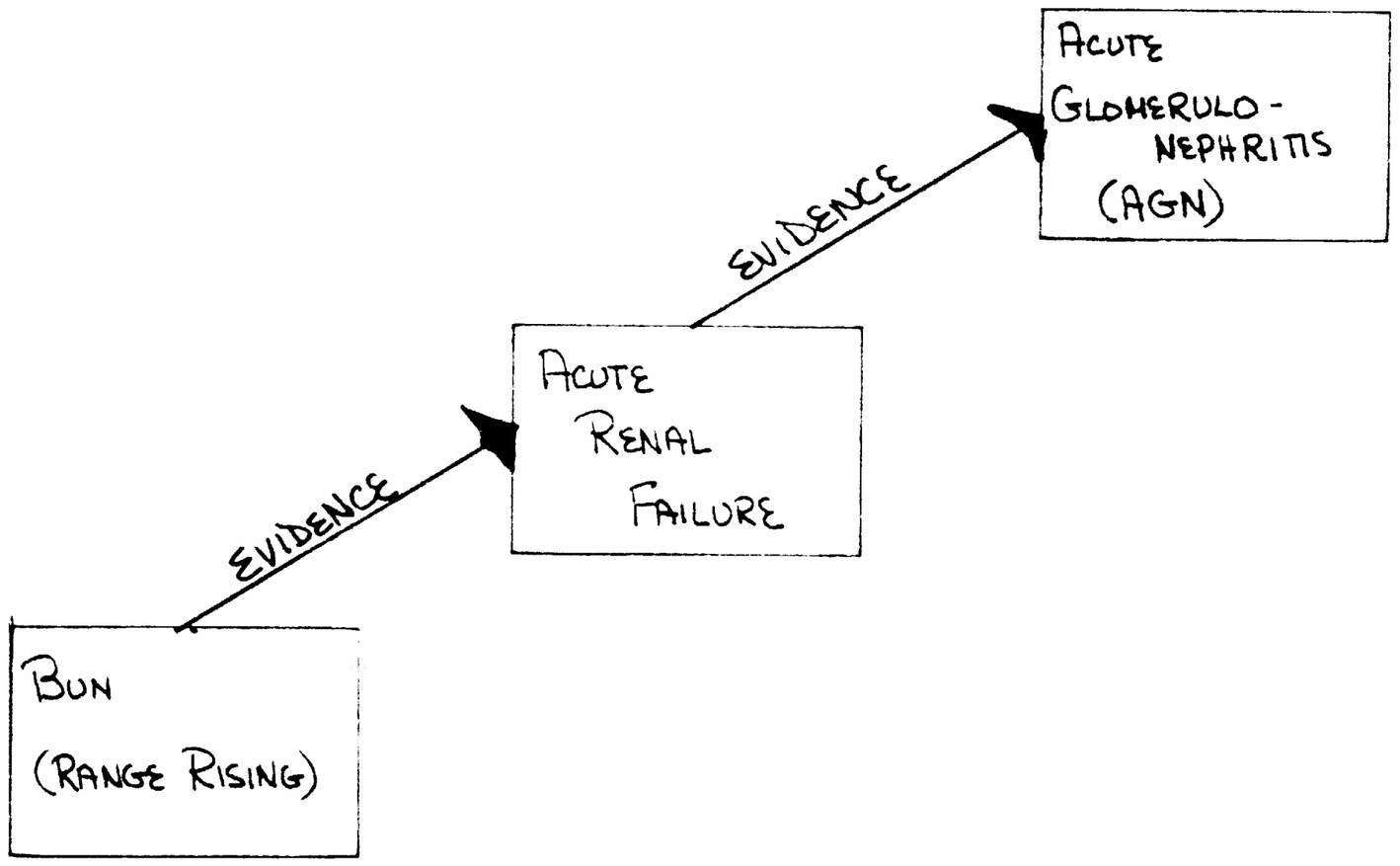


Diagram 6-2: Evidence Chain

presence of chains of EVIDENCE pointers; this means that the symptoms of a disease may not be attached directly to its elementary hypothesis, but rather to an intervening syndrome or collection of symptoms. For example, BUN (=blood urea nitrogen) (RANGE RISING) is evidence for RENAL-FAILURE ACUTE, which is in turn EVIDENCE for AGN (See Diagram 6-2). Because such structures emphasize the need for extra mechanism, it is worthwhile understanding what they contribute to the theory and its efficiency.

The clustering of the symptoms of a disease into chunks which share a common mechanism is basically a "memory hack." After I first read the description of AGN in a medical textbook, I remembered only a few scattered symptoms of the disease. After re-reading the chapter and organizing the symptoms (with the help of Gerry Sussman and Steve Pauker) into five main groupings - SODIUM-RETENTION, ACUTE-RENAL-FAILURE, GLOMERULITIS, HYPERTENSION ACUTE and STREP-INFECTION (preceding AGN by 2 to 3 weeks) - I found all the symptoms easy to recall. This intermediate-level (between symptom and actual disease) structure demonstrates how our long-term-memory (LTM) data structures exhibit "chunking." (The role of chunking in short-term-memory (STM) has already been alluded to in Chapter 3; see Chapter 7 for a more theoretical and comprehensive discussion of the chunking phenomenon.) This is not the place to propose a theory of memory, but it seems obvious that such a multi-level structure should make access to and recall of the symptoms of a disease (or any other

data) easier.

In addition, the intermediate-level structures like SODIUM-RETENTION or ACUTE-RENAL-FAILURE are useful in other contexts: other diseases, like RENAL-INFARCTION, exhibit the symptoms of ACUTE-RENAL-FAILURE, and CIRRHOSIS, for example, may cause generalized SODIUM-RETENTION. Thus, representing these structures as independent entities saves space; they don't have to be remembered separately as part of several different diseases. In addition, because they are whole sub-assemblies, they can be moved from hypothesis to hypothesis during diagnosis without being reassembled. This is one of the reasons why the price for wrong guessing is fairly low - if a hypothesis is wrong, many of its sub-hypotheses (like ACUTE-RENAL-FAILURE) and their associated symptoms can be transferred en masse to another hypothesis. When the manifestation of one of these general syndromes differs between diseases, that may be represented by an OVERRIDE assertion, as explained below in Section 6.5.2 and above in Chapter 5.

The existence of these structures also allows the generation and evaluation of a hypothesis corresponding to a misfunctioning mechanism without regard to the specific disease in which it occurs. The necessity of evaluating more general hypotheses is even clearer when we consider G-U-TUMOR and its more specific examples - KIDNEY-TUMOR, BLADDER-TUMOR etc. It is certainly better to be able to activate only G-U-TUMOR and consider the findings on that level until

a definitely discriminating symptom (such as an IVP showing a mass in a particular location) shows up rather than considering all the particular examples immediately. As mentioned above in Chapter 5, doctors often jump directly to a very specific etiology, but in cases of insufficient information, they must be able to entertain more general hypotheses. I have seen Dr. Kassirer postulate something as general as INFECTION to relate and account for a fever and the presence of white blood cells in the spinal fluid when he was unsure as to the ultimate diagnosis.

Thus, the placement of symptoms and diseases into this multi-level structure seems both intuitively and theoretically justified; one of the major chores of the global assembly phase is to put back together what has been separated by this process.

6.4 Global Assembly's Chores

There are four aspects to the job the global assembly phase must accomplish, chores which the local evaluation phase could not perform having access to the context of only one elementary hypothesis. A more complete description of each chore follows this quick list.

1. Put together several elementary hypotheses and, perhaps, unattached symptoms into a larger hypothesis using ISA, EVIDENCE, CAUSE, COMPLICATION-OF and DEVELOPS-INTO links. This

may require activating one or more new elementary hypotheses to fill in the spaces between already active ones. If so, their scores will be calculated, requiring a temporary return to the local evaluation stage.

2. Check global differential diagnosis assertions which may result in deferring an elementary hypothesis because a more likely diagnosis exists.

3. Examine the various members of CHOICE-SETS which are active with the hope of being able to accept or reject additional elementary hypotheses because of the information inherent in the structure of a CHOICE-SET.

4. Form adequate hypotheses which account for all the abnormalities present. This chore requires using the designation ULTIMATE-ETIOLOGY and forms complex hypotheses which contain more than one <u>independent</u> component. The results are disease-centered hypotheses, rather than hypotheses which consist of one disease; in the protocol, for example, the final diagnoses were LGN- and FGN-CENTERED. This chore can become extremely complex, as it theoretically involves discovering the best partition of findings into separate elementary hypotheses, a problem which I have far from solved. The relationship between this chore and the disposing stage of processing explained in Chapter 3 will be examined as well.

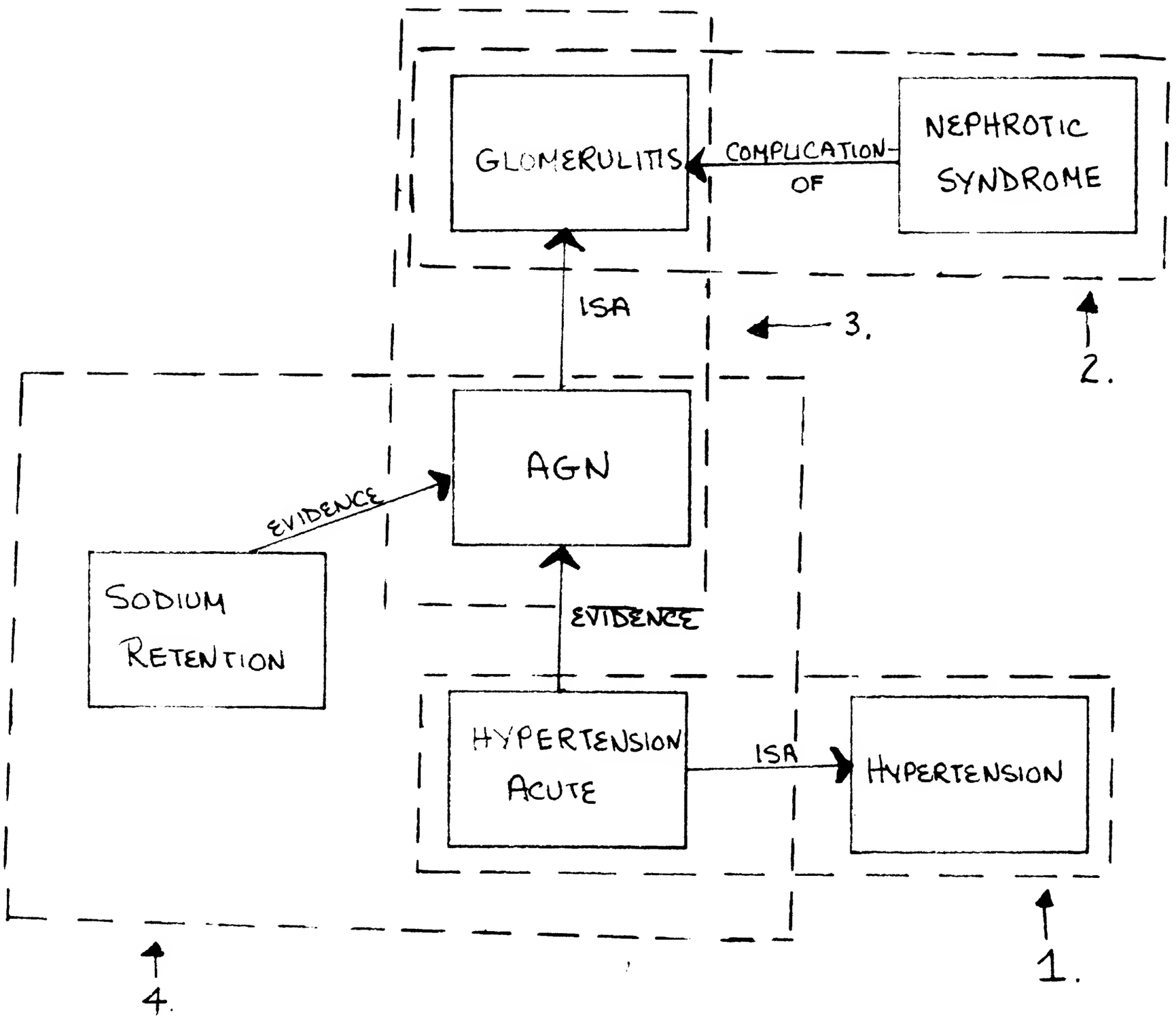
The most complicated and important of these chores are 1 and 4 - the

formation of coherent and adequate hypotheses.

6.5 Chore 1: Forming Coherent Hypotheses

A coherent hypothesis consists of two or more elementary hypotheses joined by "coherence links" which include ISA, CAUSE, COMPLICATION-OF, DEVELOPS-INTO and EVIDENCE links. Coherent hypotheses are constructed out of already-active hypotheses and, perhaps, some inactive ones as well, which are activated in the course of performing the chore. The local evaluation function of the newly-activated elementary hypothesis may, in fact, be composed of the local evaluation functions of other elementary hypotheses in the structure, suitably combined. This is particularly the case in EVIDENCE-chained hypotheses, discussed in Section 6.5.2.

The construction of coherent hypotheses is a repeatable process in that any of the nodes in a coherent hypothesis may be an elementary hypothesis which is itself already part of a coherent hypothesis. In such a case, the two coherent hypotheses become part of a larger structure; Diagram 6-3 is an example of a complex coherent hypothesis which contains four separate coherent hypotheses, as indicated by the dotted rectangles. The descriptions of coherent hypothesis structures below may be regarded as "templates" which are placed on the patient's data structure; present findings, active hypotheses and necessary links are labelled more darkly in the



THERE ARE FOUR SEPARATE COHERENT HYPOTHESES IN THE DIAGRAM, AS INDICATED BY THE NUMBERS.

Diagram 6-3: COMPLEX COHERENT HYPOTHESIS

diagrams which follow and are the structures which must match. The action consists of joining the matched components together, along with newly-activated hypotheses which are outlined more lightly. For clarity, the actions taken when the template fits are expressed in words in the diagrams, as well as being implicit in the drawn structure.

6.5.1 ISA-connected hypotheses

ISA-connected hypotheses are the simplest type of coherent hypothesis dealt with by this stage. They consist of two elementary hypotheses, one of which is an example of the other. In the protocol, for example, there were many examples of such hypotheses: AGN ISA GLOMERULITIS, PYELONEPHRITIS ISA G-U-TRACT-BLEEDING etc. Calculating the score of the more specific disease, which is considered the score of the entire composite hypothesis, requires taking into account those symptoms attached to the category node, as well as those linked directly to the example. A concrete example of the relationship between symptoms and scores of the two hypotheses using STREP-INFECTION is explained in detail in Chapter 4. Consider, as another example, G-U-TUMOR and KIDNEY-TUMOR. As illustrated in Diagram 6-4, the two together form a coherent hypothesis whose score is that of the more specific hypothesis, in which those symptoms relevant to the more general category are taken into account.

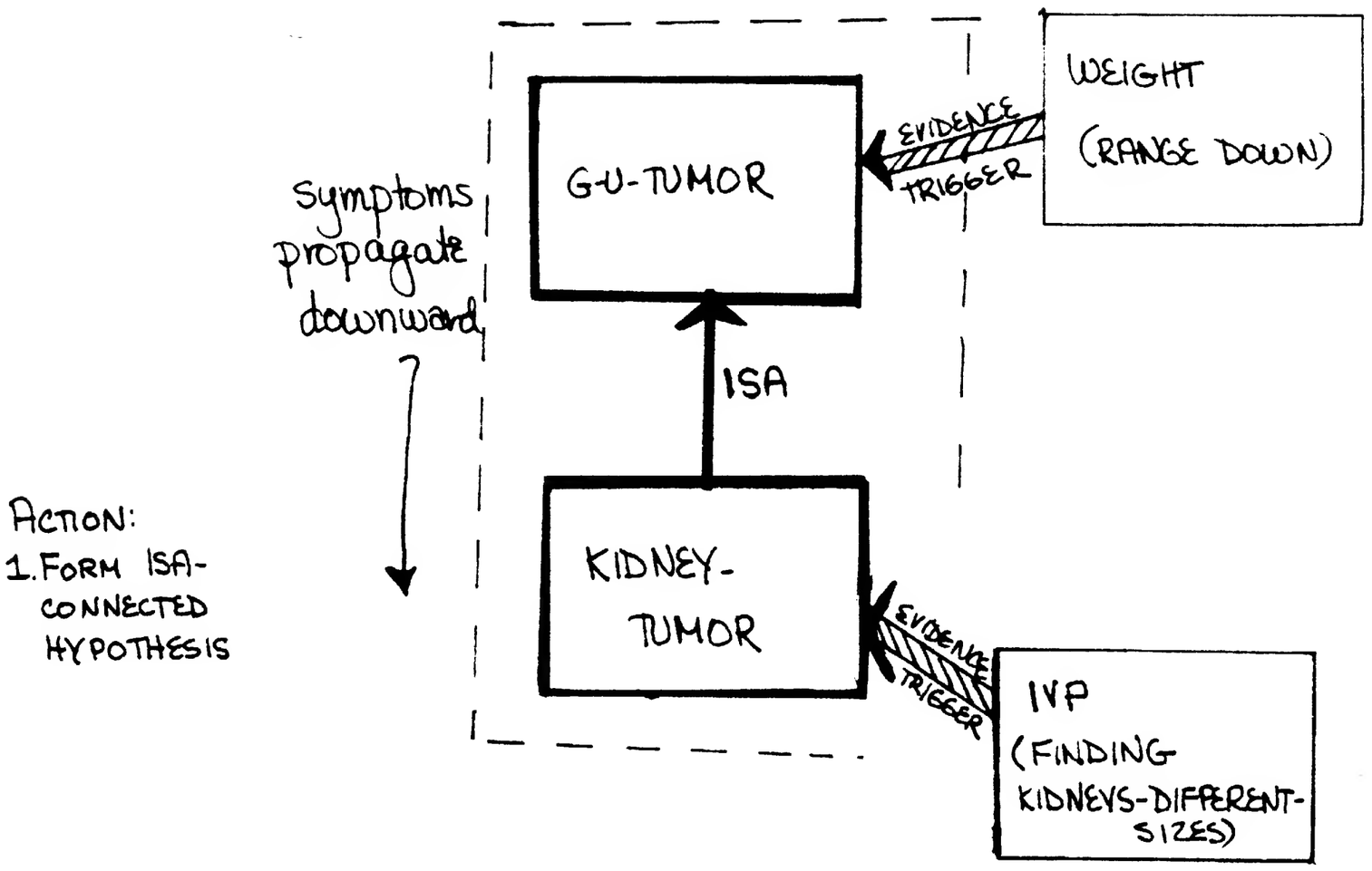


Diagram 6-4: ISA-CONNECTED TEMPLATE

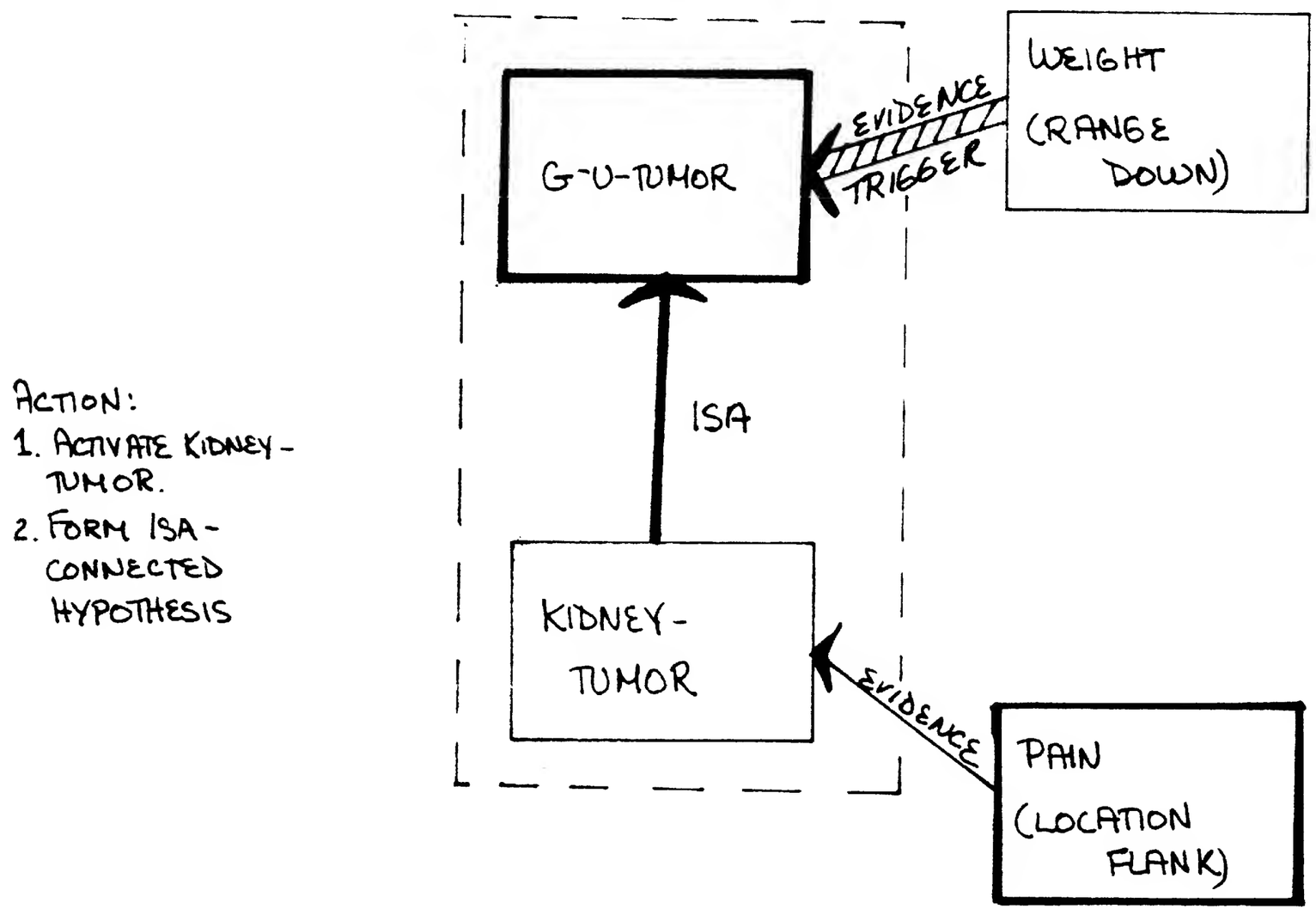


Diagram 6-5: ISA-CONNECTED TEMPLATE REQUIRING ACTIVATION OF ONE ELEMENTARY HYPOTHESIS

In order for an ISA-connected hypothesis to be formed, of course, the component elementary hypotheses must be active, as illustrated in Diagram 6-4. This happens in the local evaluation phase, since it is necessary to evaluate the category of a disease in order to evaluate the specific disease hypothesis at all. Even though it is clear that the relationship between the elementary hypotheses corresponding to a category and an example must be discovered before or during the local evaluation stage, I have included the description of this type of coherent hypothesis here because it fits conceptually with the following kinds of coherent hypotheses.

A second kind of ISA-connected coherent hypothesis requires the activation of an elementary hypothesis before it can be formed. When a category is active and a finding is added which is relevant to one of the CHOICE-SET members, but does not trigger it by itself, there is enough evidence to activate that CHOICE-SET member. Such a situation is exemplified in Diagram 6-5. Such a finding may, in addition, be a SUFFICIENT CHOOSER in that, in the context of the CHOICE-SET, it can be accounted for by only one elementary hypothesis. As explained in Chapter 4, if a category is accepted, a SUFFICIENT CHOOSER pointing to one of its examples is enough to accept that specific hypothesis. In both cases, the CHOICE-SET is acting as a smaller-than-global context in which the particular finding is significant; in a global context it would not activate (or accept) the relevant CHOICE-SET member.

6.5.2 EVIDENCE-chained hypotheses

As explained above in Section 6.3, symptoms are often not connected directly to their diseases, but to intermediately-general pathological states. It is thus important to be able to reconnect the symptoms to the actual disease; this is done via EVIDENCE-chained hypotheses. They are formed when two or more active elementary hypotheses have EVIDENCE chains which intersect at a single etiology. For example, if SODIUM-RETENTION and ACUTE-RENAL-FAILURE were both active, we would want to unify them into a hypothesis which postulated AGN. See Diagram 6-6 for an illustration of a template for this type of hypothesis; I shall call the disease hypothesis at which the chains intersect the "center" in what follows.

The relevant symptoms of a disease-hypothesis formed in this way are all those relevant to any of the intermediate structures, as well as any which may be attached directly to the center itself. The strengths of EVIDENCE and EXPECTATION between each symptom and the elementary hypothesis to which it is directly connected may be modified by the EVIDENCE or EXPECTATION pointer between that hypothesis and the center. In determining the contribution of relevant symptoms to the center's score, we multiply the EVIDENCE and EXPECTATION strengths of symptom to intermediate hypothesis by the

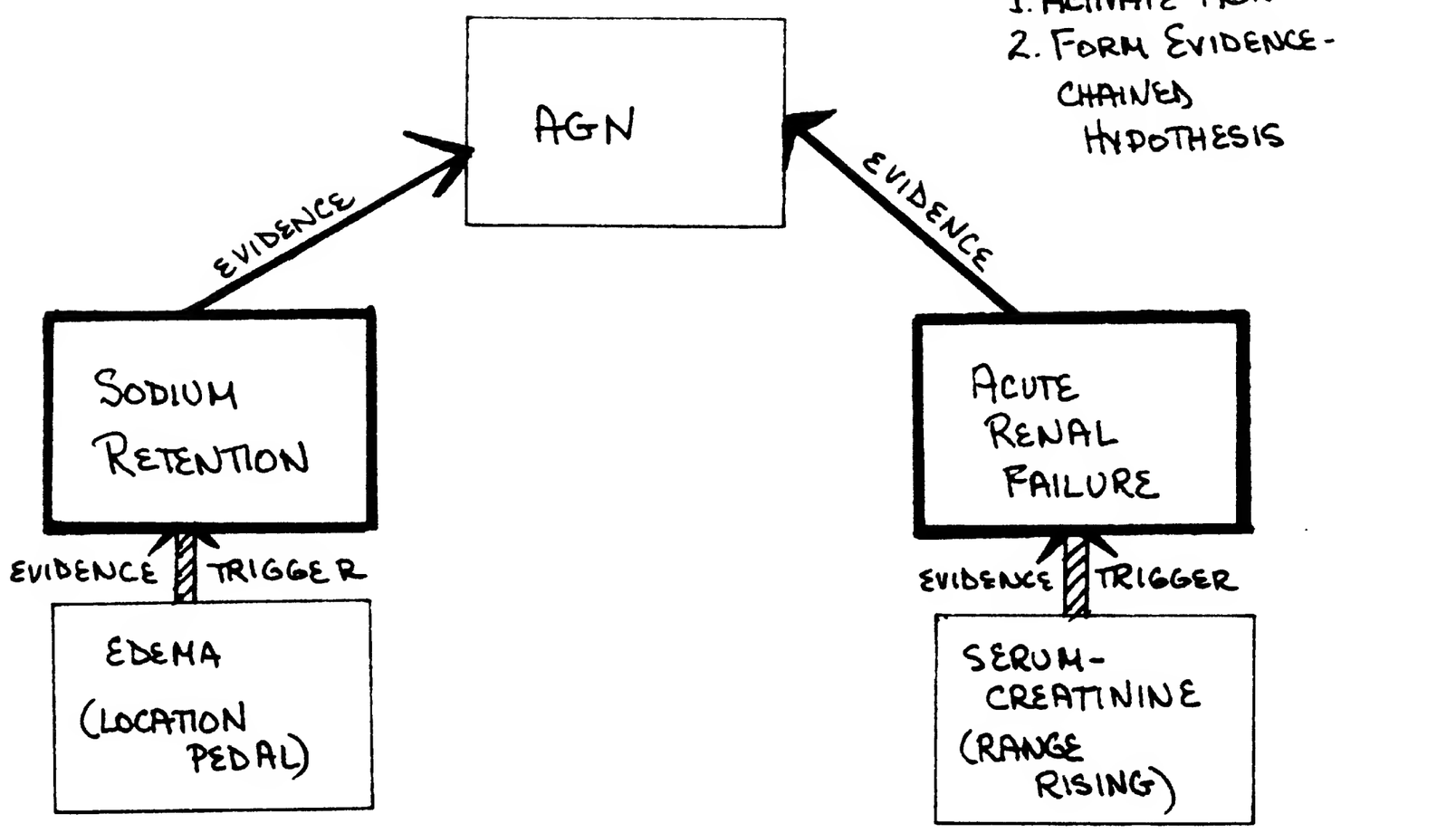
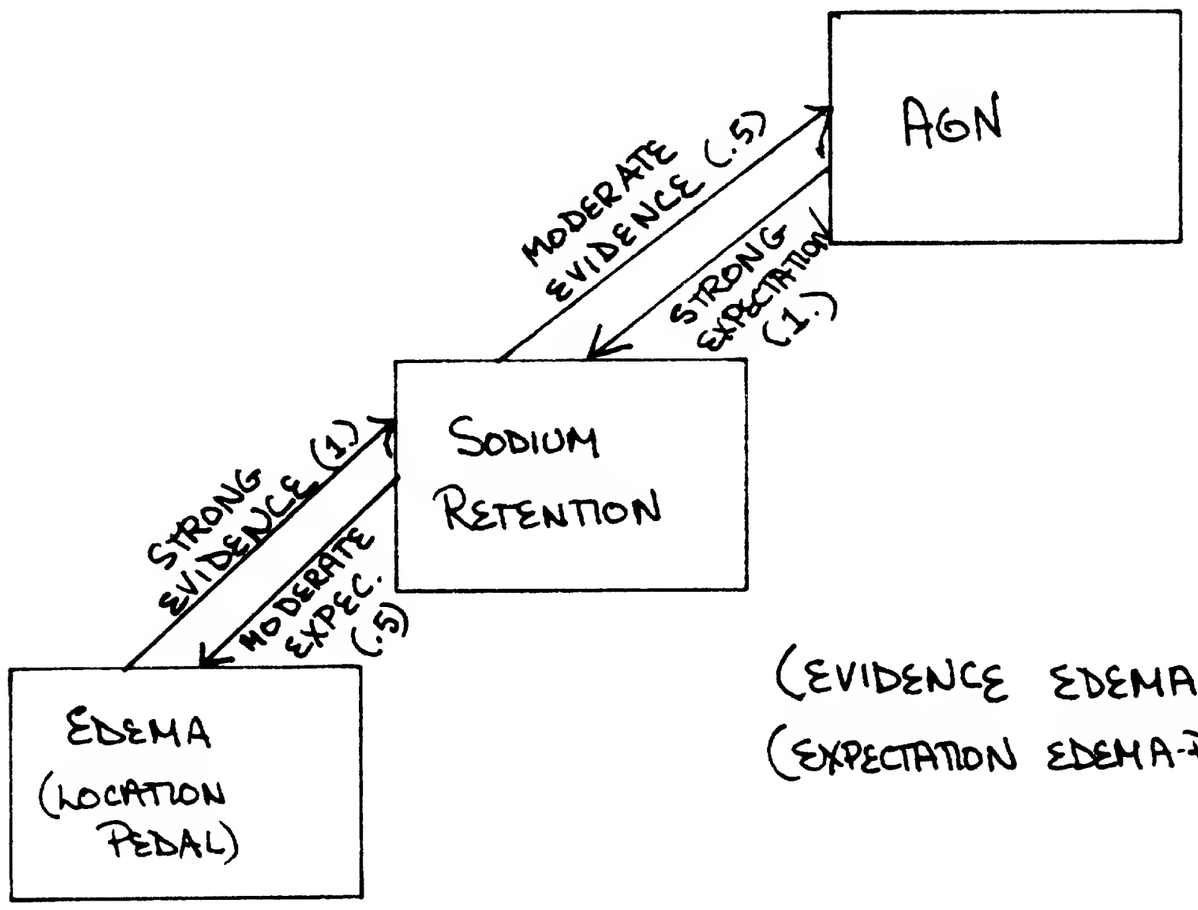


Diagram 6-6: EVIDENCE-chained template



(EVIDENCE EDEMA-PEDAL AGN 1X.5=.5)
(EXPECTATION EDEMA-PEDAL AGN .5X1 = .5)

Diagram 6-7: COMBINING EVIDENCE and EXPECTATION STRENGTHS ALONG CHAINS

corresponding strength (i.e. EVIDENCE for EVIDENCE, EXPECTATION for EXPECTATION) between intermediate hypothesis and center. Diagram 6-7 should make this interplay clearer. SUFFICIENT EVIDENCE and NECESSARY EXPECTATIONS combine in the obvious way.

As a slight variation on the template drawn in Diagram 6-6, we can consider the case of an elementary hypothesis and a non-trigger finding interacting to activate another elementary hypothesis for which they are both evidence. This would be the case with ACUTE-RENAL-FAILURE and YOUNG, activating AGN as illustrated in Diagram 6-8. It is also possible that a structure like that in Diagram 6-9, in which the relevant finding or an active hypothesis is separated from the ultimate center of the hypothesis by two links may also cause the formation of a coherent hypothesis, but I do not have any relevant examples.

This last speculation brings up an important point: does the distance between nodes in a template for a complex hypothesis affect its coherence? The structure suggested in Diagram 6-9 requires the activation of two extra elementary hypotheses and the finding on the left is separated from the center by two links. The other proposed templates only require the activation of one extra hypothesis and the template-matching findings and hypotheses are only separated from the center by one link. In EVIDENCE-chained hypotheses, it seems that there should be no limit on the number of intervening links, as the intermediate structures exist for ease of memory and conceptualization

and don't really represent epistemological or medical "distance." Symptoms of SODIUM-RETENTION <u>are</u>, after all, symptoms of AGN, no matter how many intervening elementary hypotheses there are. For now, then, I will not postulate any limit on the number of EVIDENCE links between a present finding or active hypothesis and the eventual center of a template. Conceptually, we should think of a chain of EVIDENCE pointers as a single one. This collapsing of EVIDENCE links also occurs in the next type of coherent hypothesis considered, that connected by CAUSE, COMPLICATION-OF or DEVELOPS-INTO links.

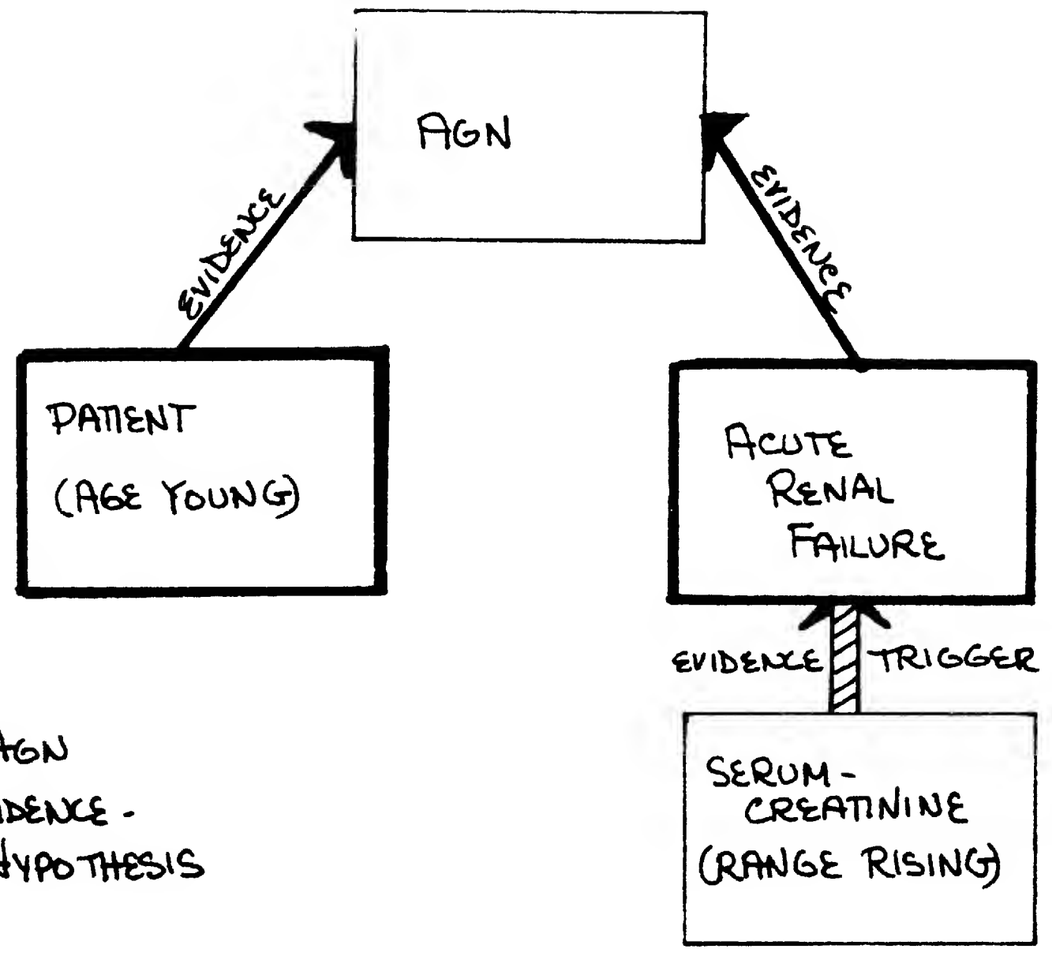
There is a processing issue here as well. When some general procedure is searching for template matches, any straightforward implementation will have trouble with a combinatorily-expanding search if we allow any number of intermediate links between finding and elementary hypothesis. How can we avoid this? Perhaps another example of compilation of global knowledge is to be found here; there may be a special piece of information in one of the sub-hypotheses which checks for the other sub-hypothesis and, if conditions are right, activates the "center." For example, the SODIUM-RETENTION elementary hypothesis might specifically check to see if ACUTE-RENAL-FAILURE is active and, if so, activate AGN.

One thing which can prevent the formation of EVIDENCE-linked hypotheses is an OVERRIDE assertion, which overrides the chain of EVIDENCE pointers by asserting that a symptom is RARE in a particular disease, even though it is good evidence for the intervening

pathological state. The canonical example of this situation, introduced as the "X" phenomenon in Chapter 5, is the SODIUM-RETENTION hypothesis. Both facial edema (fluid retention in the face) and ascites (fluid in the gut) are symptoms of SODIUM-RETENTION; both AGN and CIRRHOSIS (liver malfunction) exhibit SODIUM-RETENTION. The location of fluid retention differs in the two diseases, however, so the formation of a hypothesis which contains facial edema as evidence for CIRRHOSIS or ASCITES as evidence for AGN through SODIUM-RETENTION is precluded by an OVERRIDE condition. Diagram 5-1 is repeated as 6-10 for reference.

The use of an OVERRIDE assertion containing the designation RARE to brand a hypothesis <u>incoherent</u> corresponds to the use of the designation VERY-RARE in a priori probabilities to <u>reject</u> an elementary hypothesis. Both are examples of a heuristic which keeps the number of active hypotheses low by not considering at all those which are highly improbable. Of course, it is <u>possible</u> that AGN might cause ASCITES or CIRRHOSIS might cause EDEMA (LOCATION PEDAL) and rejecting such a hypothesis out of hand may lead a doctor or diagnostic system through a long and tortuous search for another explanation. Here is another example of a place where backtracking from considering only coherent hypotheses to allowing some incoherent ones is necessary.

The formation of all coherent hypotheses besides ISA-connected ones is sensitive to the time-information contained in the data



ACTION
1. ACTIVATE AGN
2. FORM EVIDENCE-
CHAINED HYPOTHESIS

Diagram 6-8: EVIDENCE-CHAINED TEMPLATE

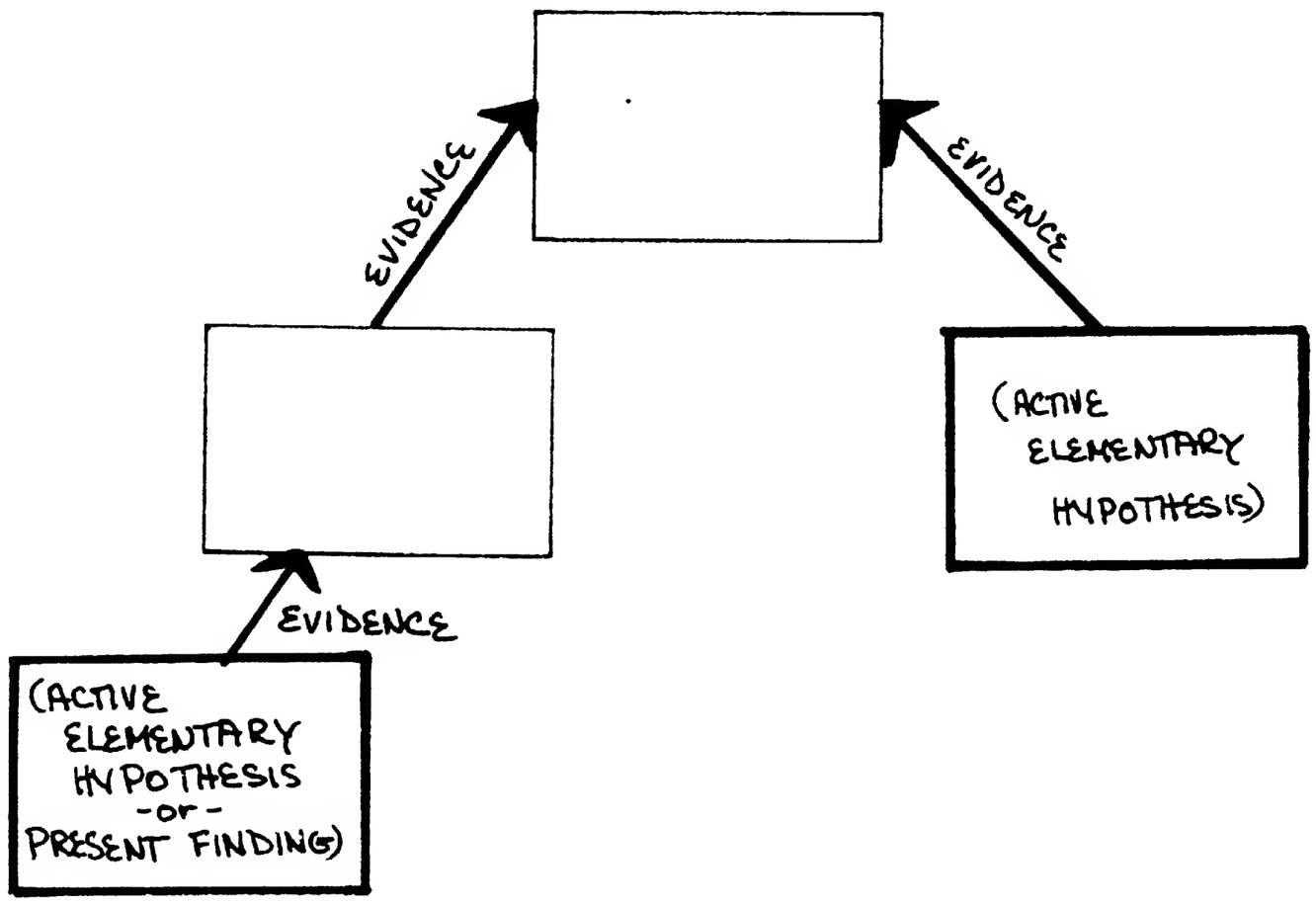


Diagram 6-9: ANOTHER EVIDENCE-CHAINED TEMPLATE?

network. If a BEFORE time-relationship is specified in the data, it must be satisfied by the time-instantiated elementary hypotheses which form the coherent hypothesis.

Often the above coherent hypothesis mechanism may be unnecessary, as a disease hypothesis may be triggered directly by the findings, rather than the more general hypotheses. This type of triggering indicates the compiled nature of the doctor's knowledge. (See Section 5.4) In the AGN example pictured above in Diagram 6-6. EDEMA (LOCATION PEDAL) and SERUM-CREATININE (RANGE RISING) could have been a multiple trigger for AGN, representing a compiled version of the global assembly mechanism proposed here. When a disease hypothesis such as AGN is activated, all of the pathological states which are evidence for it are also activated, since computing a score for AGN requires computing scores for each of the subgroups of symptoms. It is by this activation of sub-hypotheses and propagation of finding contributions along EVIDENCE chains that CGN was ruled out in the protocol. (BUN (RANGE HIGH)) is a NECESSARY EXPECTATION in CHRONIC-RENAL-FAILURE; CHRONIC-RENAL-FAILURE is in turn a NECESSARY EXPECTATION in CGN. When CGN was activated, so was CHRONIC-RENAL-FAILURE and the appropriate combinations of EVIDENCE strengths figured out. Thus, (BUN (RANGE HIGH)) was a NECESSARY EXPECTATION in CGN and its absence ruled out CGN.

6.5.3 CAUSE, COMPLICATION-OF and DEVELOPS-INTO hypotheses

These three types of coherent hypotheses are treated equivalently. As noted above in Chapter 3, CAUSE and COMPLICATION-OF links are epistemologically similar, differentiated mainly by how well the causal connection between the two entities is understood. DEVELOPS-INTO differs only in its implicit assumption of some time-dependence between the connected diseases. All of the three are sensitive to time relationships explicitly stated in the data network. I will illustrate the general form of coherent hypotheses containing these relationships and their manner of formation only once, rather than separately for each of the three specific links.

Forming a coherent hypothesis of this sort may not involve activating any new elementary hypotheses at all. If two active hypotheses are connected by a CAUSE, COMPLICATION-OF or DEVELOPS-INTO link, they may be joined into one composite hypothesis. An example of this situation is contained in Diagram 6-11, using the template conventions developed above. Usually GLOMERULITIS will be a part of another coherent hypothesis: an ISA-connected one whose other component is a specific disease such as AGN or FGN. (see Diagram 6-3 for an example of these two types of coherent hypotheses combined.)

More interesting is the case where a new elementary hypothesis must be activated. A template for this situation is contained in Diagram 6-12. CLOTTING-DISORDER is activated in order to provide the

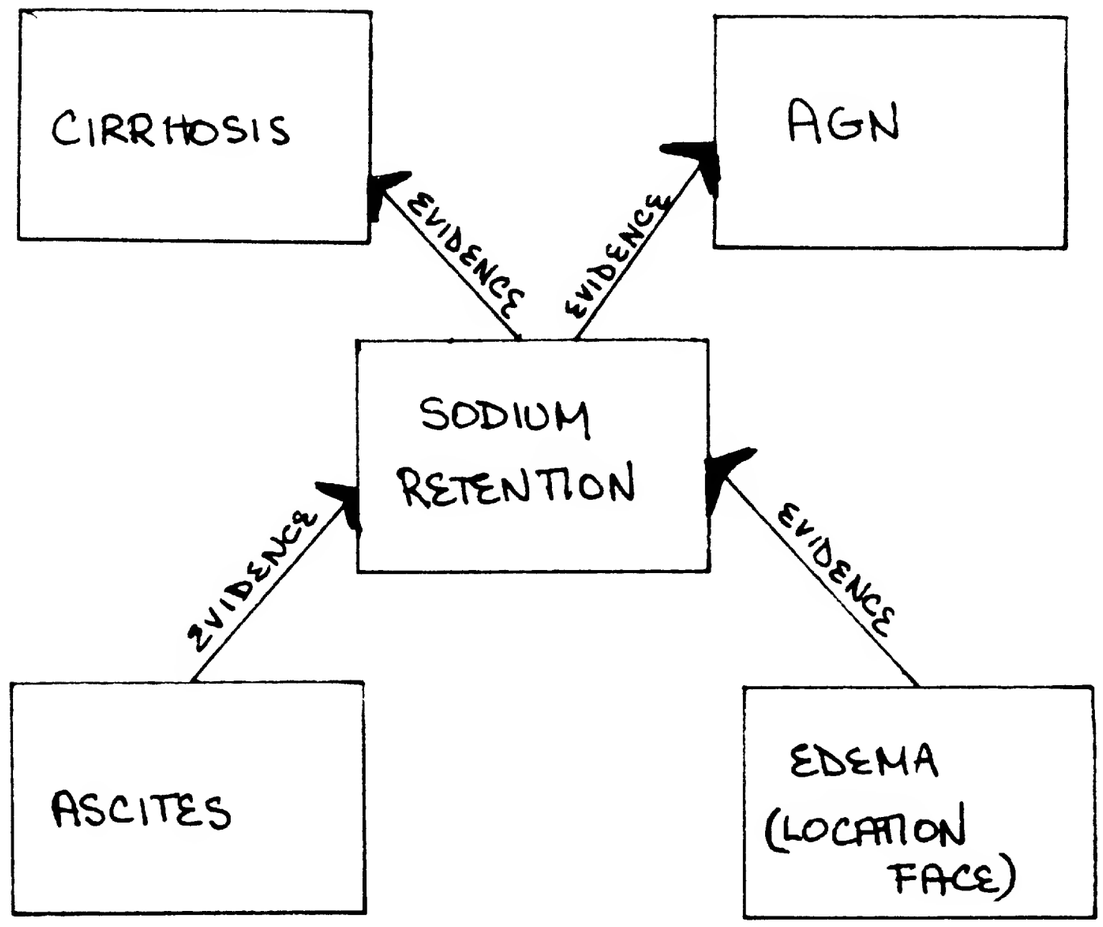
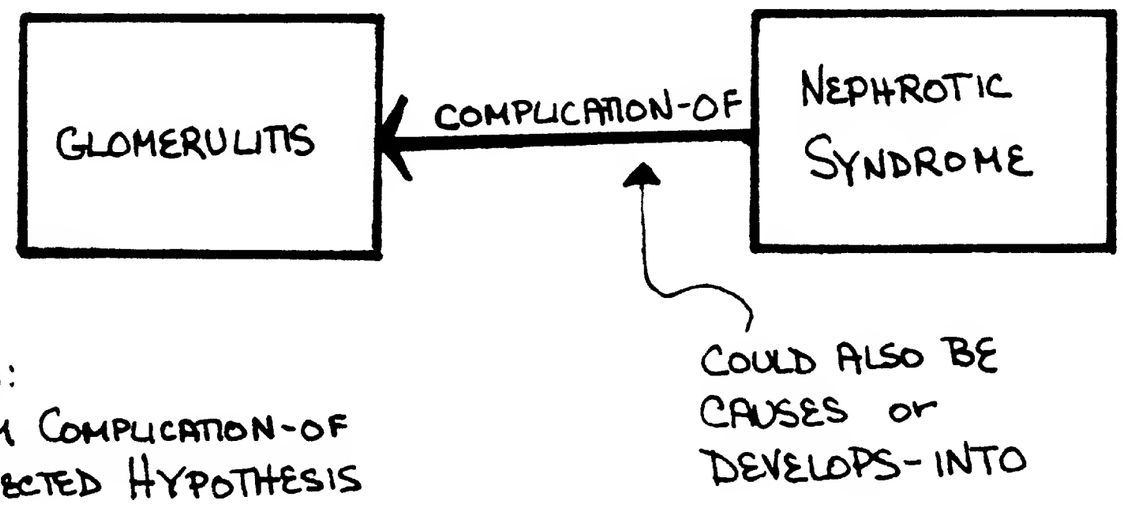


Diagram 6-10 The "X" Phenomenon

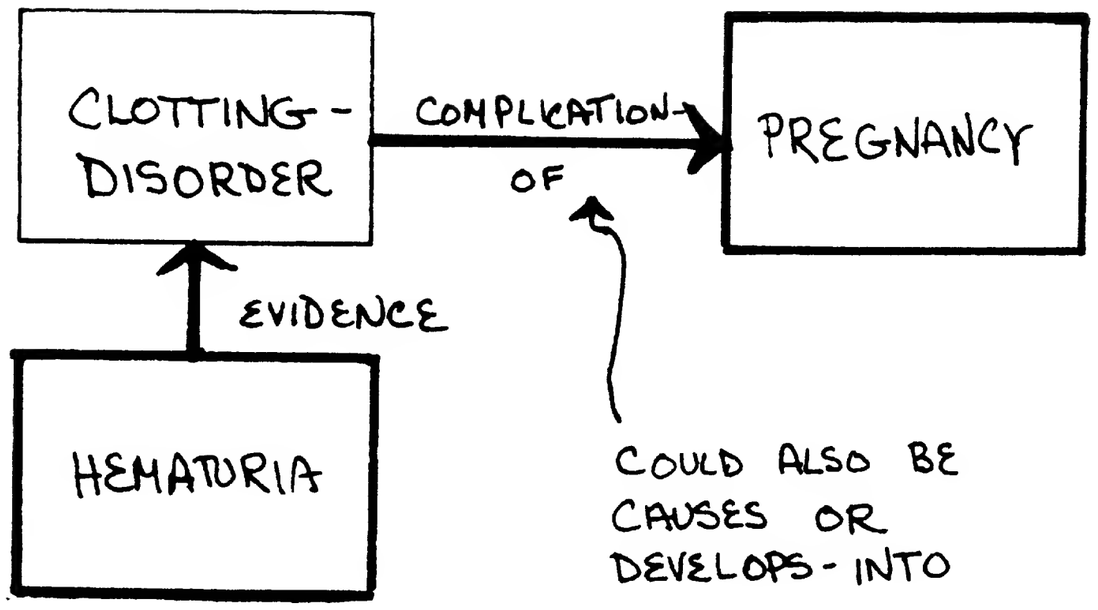


Action:
1. Form Complication-of Connected Hypothesis

Diagram 6-11 General Cause, Complication-of, Develops-into Template

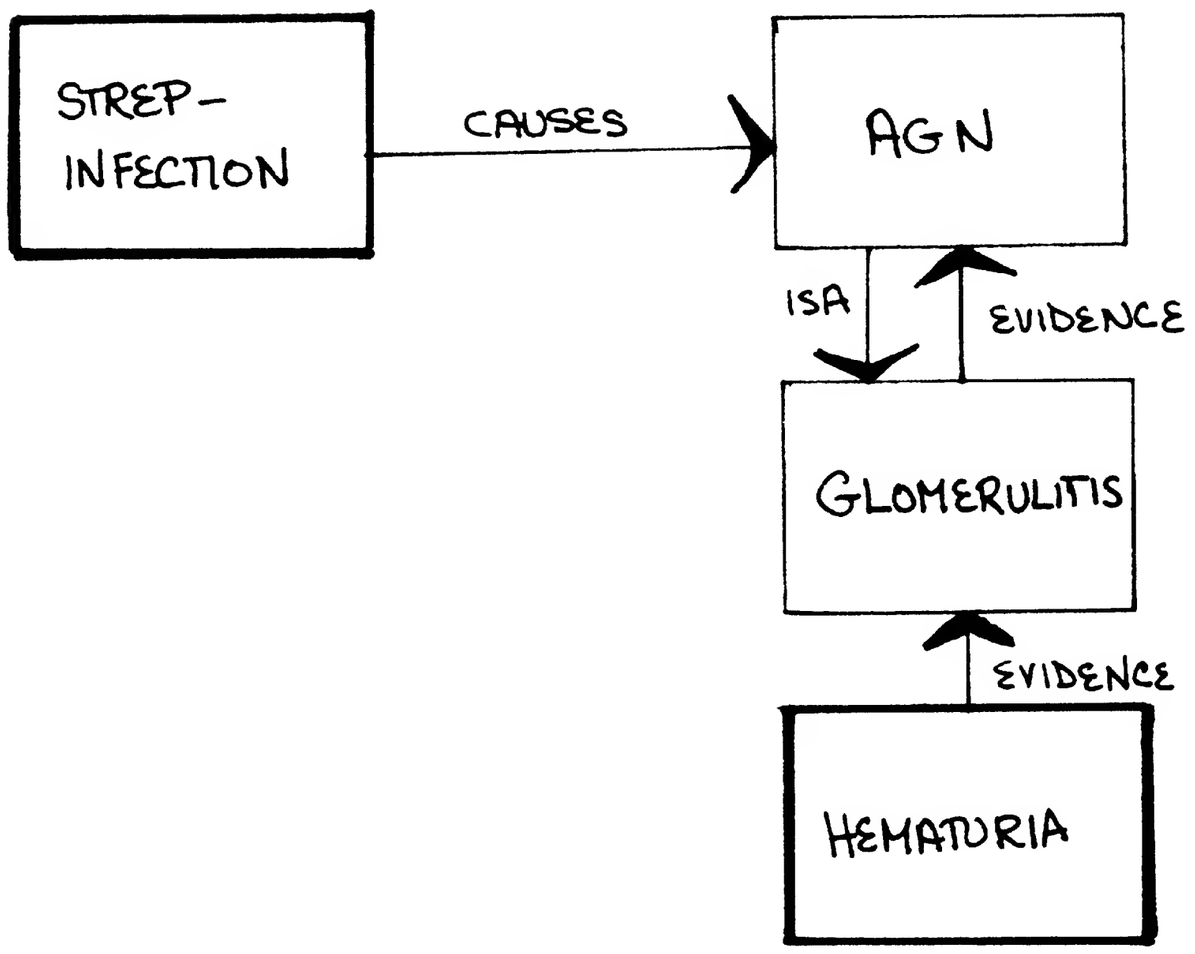
link between HEMATURIA and PREGNANCY. In words, we can activate an intermediate hypothesis which has one (non-trigger) finding present and is also connected by a CAUSE, COMPLICATION-OF or DEVELOPS-INTO link to an active hypothesis. The newly-activated hypothesis provides a link between the two. The EVIDENCE link may be replaced by a chain of EVIDENCE pointers or by a CAUSE, COMPLICATION-OF or DEVELOPS-INTO link to another active hypothesis to form other templates for this type of coherent hypothesis.

We may ask the same question about the number of links in a template as was posed above in talking about EVIDENCE-chained hypotheses: can we expand these templates so that the present findings and active hypotheses are separated by more pointers? In contrast to the EVIDENCE case, indefinitely long chains of CAUSE, COMPLICATION-OF or DEVELOPS-INTO pointers which have no supportive findings for intermediate hypothesized diseases do not seem to be acceptable templates for coherent hypotheses. Suppose, for example, we knew a patient had HEMATURIA and had had a STREP-INFECTION several weeks earlier. Two possible complex hypotheses for this situation are illustrated in Diagram 6-13. The relevant time-relationships would have to be checked in both, but the first is clearly preferable to the second because it must hypothesize fewer intermediate stages for which little evidence exists. The upper structure fits the definition of a template for a CAUSE-connected hypothesis, since we have allowed chains of EVIDENCE links anywhere an EVIDENCE link is indicated in a



ACTION:
1. ACTIVATE CLOTTING-DISORDER
2. FORM COMPLICATION-OF CONNECTED HYPOTHESIS

Diagram 6-12 GENERAL CAUSES, COMPLICATION-OF, DEVELOPS-INTO TEMPLATE



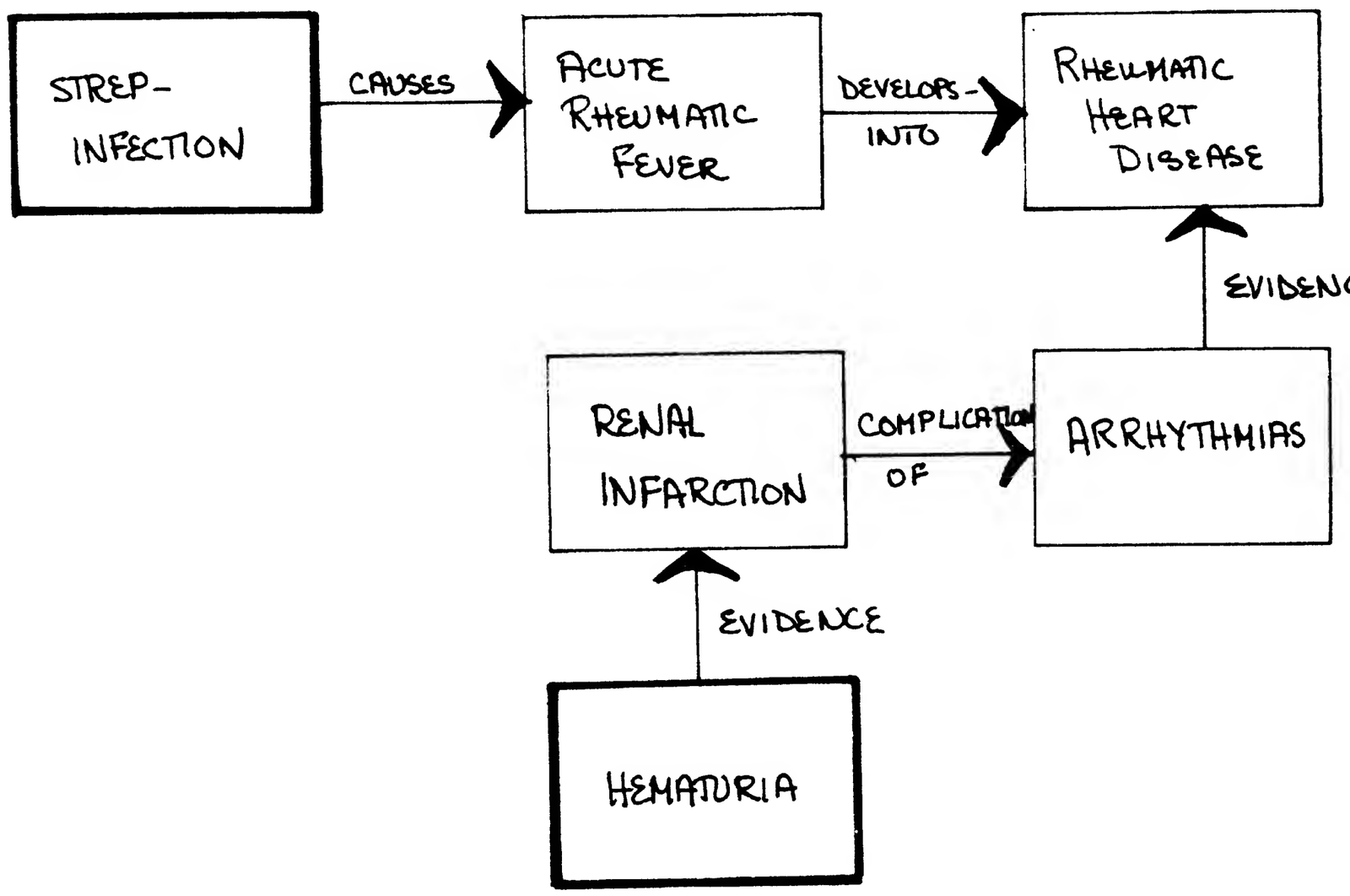


Diagram 6-13: Two Possible Complex Hypotheses
Connecting Hematuria and Strep-Infection

template. The lower structure, however, has too many intermediate steps between the solid data; the system (and, I feel, doctors) would not accept a hypothesis such as this without further evidence. Putting a limit on the number of intermediate links in a template eliminates the possibility of discovering long tortuous paths between any two entities (findings or hypotheses) in the patient's condition. Again, it is a heuristic for limiting the number of possibilities which may result in overlooking a hypothesis which finally turns out to be the correct diagnosis.

CAUSE, COMPLICATION-OF and DEVELOPS-INTO linked hypotheses are all methods of dealing with findings which are unaccounted-for by already-active hypotheses. For example, in the protocol, LGN, FGN and PCKD were being considered when HYPERTENSION CHRONIC was introduced (actually triggered and substantiated by ANTIHYPERTENSIVE-DRUGS (STATUS TAKEN)); LGN couldn't account for the hypertension, but a coherent DEVELOPS-INTO linked hypothesis made the connection through CGN, as illustrated in Diagram 6-14.

6.5.4 EXCUSE hypotheses

The last kind of coherent hypothesis is different from the others in that it is really formed at the local evaluation stage; it contains two elementary hypotheses, however, and so should be classified along with the other coherent hypotheses in this section.

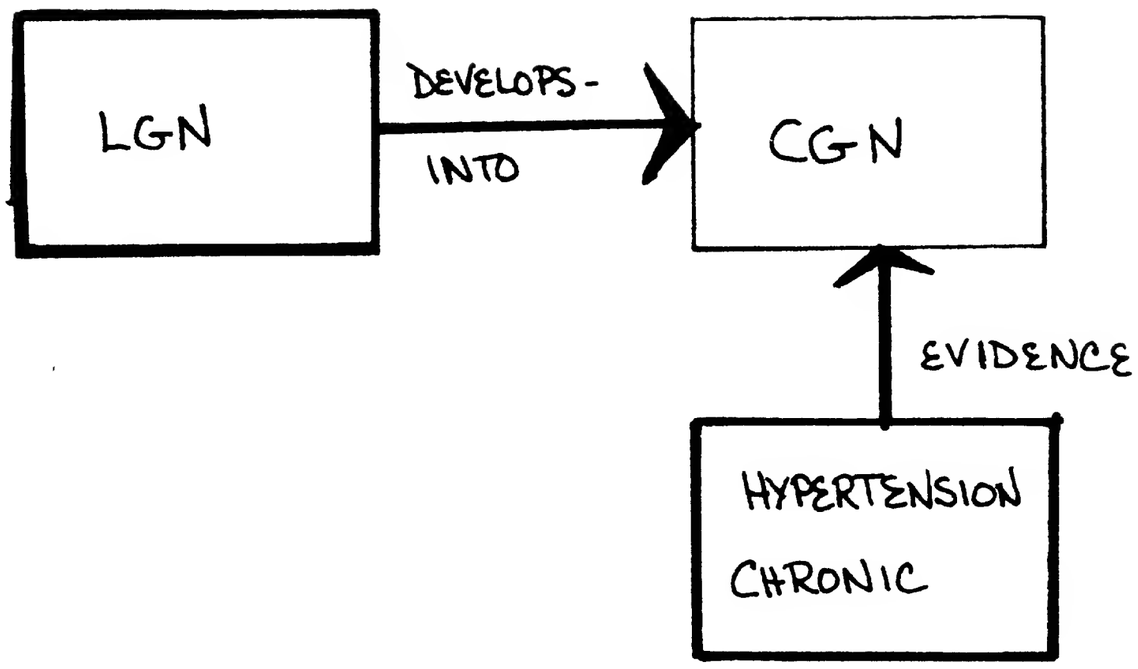
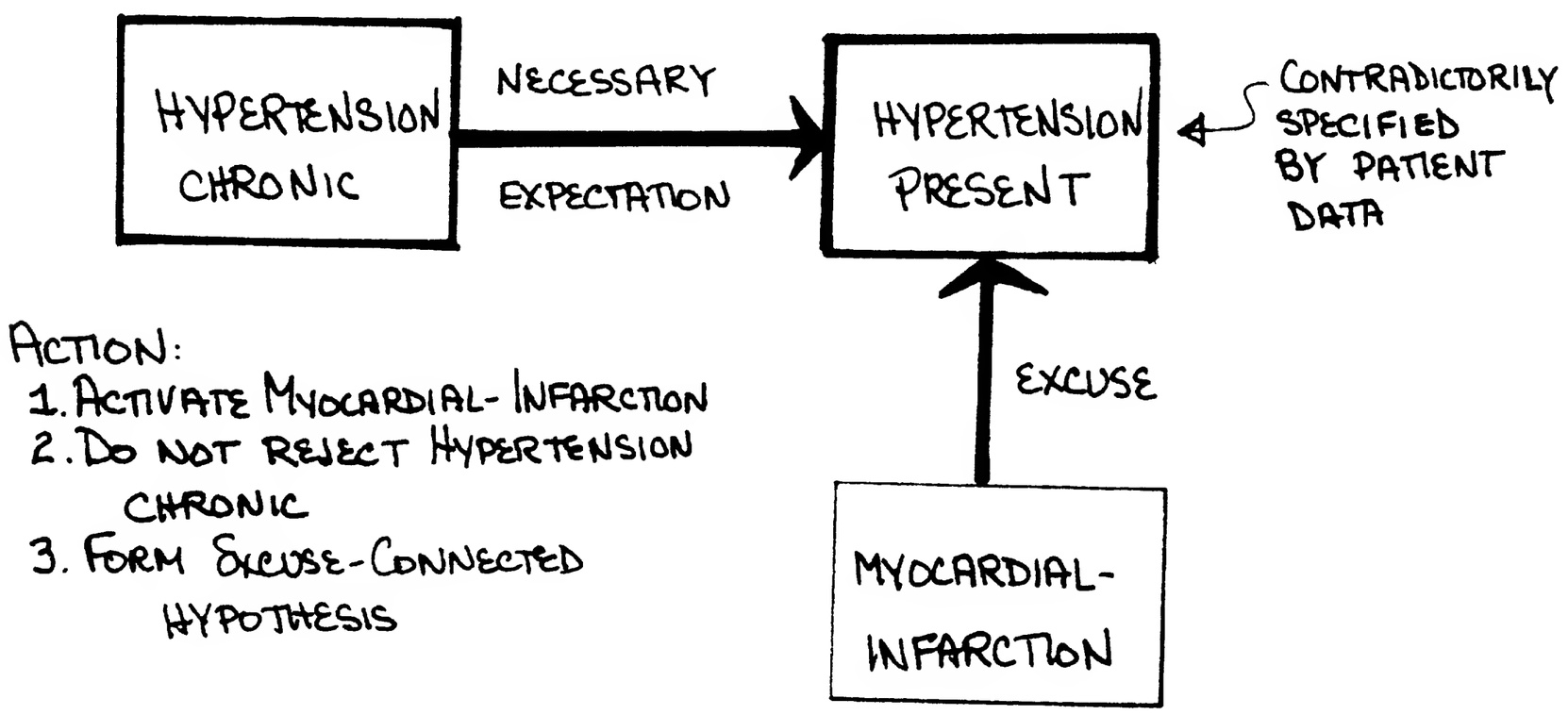


Diagram 6-14: Activation of CGN



Action:
1. Activate Myocardial-Infarction
2. Do not reject Hypertension
   chronic
3. Form Excuse-Connected
   Hypothesis

Diagram 6-15: Excuse-connected hypothesis template

While the hypotheses described in the immediately-preceding section provided methods for dealing with unaccounted-for findings, EXCUSE hypotheses provide a way to deal with violated expectations - findings expected in a disease but absent - without giving up on the hypothesis. As explained in Chapter 5, certain findings or elementary hypotheses may exist as EXCUSES for the absence of others. The example quoted there was HEMATURIA GROSS acting as an EXCUSE for the absence of RED-BLOOD-CELL-CASTS in GLOMERULITIS. Sometimes an EXCUSE is itself an elementary hypothesis whose presence must be substantiated by further evidence. Take, for example, HYPERTENSION PRESENT, which is a NECESSARY EXPECTATION in HYPERTENSION CHRONIC; a MYOCARDIAL-INFARCTION (heart attack) can act as an EXCUSE for its absence. In that case, the coherent hypothesis illustrated in Diagram 6-15 is formed; the general template is clear from that figure. The excuse is the new hypothesis which is activated by the discrepancy between expectation and reality.

6.5.5 Multiple Triggers, Viewed Again

I have mentioned several times that the formation of coherent hypotheses which activate new elementary hypotheses is similar to the mechanism of multiple triggers explained in Chapter 5. For example, we may consider HEMATURIA and PREGNANCY to be a multiple trigger for CLOTTING-DISORDER, or we may consider CLOTTING-DISORDER to be

activated by the global assembly process searching for COMPLICATION-OF connected hypotheses. What are the differences between these two conceptualizations?

Following one of the main themes of this thesis, we may view multiple triggers as a local compilation of the general knowledge used in the global assembly stage. This compiled knowledge would be contained in a type of hash table which associates with combinations of findings the elementary hypotheses which they activate. The search procedure for coherent hypotheses involved in the global assembly stage may take a while; the pre-compiled knowledge contained in a multiple trigger is more efficient.

On the other hand, the templates used in the global assembly stage make the epistemological connections between entities clearer; knowing HEMATURIA and PREGNANCY trigger CLOTTING-DISORDER says nothing about the connections between them in the data network, while the global assembly templates make clear the fact that the crucial connection is COMPLICATION-OF. However, basing the coherence of hypotheses completely on the "abstract" form of the links contained in them forces us to treat every configuration which fits a template uniformly; any time we find such a configuration we must activate the intervening hypotheses. Multiple triggers, on the other hand, are the epitome of non-uniformity. As I have described them, they are simply special assertions completely specific to the symptoms and disease they relate and without implication about how similar structures might

be treated.

Both approaches are clearly necessary in a complete system. Perhaps the uniform coherent hypothesis approach represents an earlier stage in an expert's development; as he or she becomes more expert, the information is compiled into less uniform, more efficient structures. Perhaps the multiple triggers have attached to them an EXPLANATION property which makes explicit the relational structure from which they are derived. At all stages of development, the original information embodied in the data network must still be used for explanation and debugging, as suggested in Chapter 5.

6.6 Chore 2: Global Differential Diagnosis

The global stage must also make use of global differential diagnoses which indicate which of two possible diagnoses is more likely given a certain combination of symptoms. The most concrete example of this possibility in this data is the differential diagnosis between GLOMERULITIS and G-U-TRACT-BLEEDING. As indicated in Chapter 5 (section 5.2.1), the comparative severities of HEMATURIA and PROTEINURIA suggest the two possible diagnoses differentially. Specific combinations of severities rule out each possibility and are used in local evaluation to reject elementary hypotheses, for example:

(PRECLUDES (AND (HEMATURIA GROSS) (PROTEINURIA LIGHT)) GLOMERULITIS)

Other information is expressed only comparatively, for use by the

global stage, for example:

(MORE-LIKELY G-U-TRACT-BLEEDING THAN GLOMERULITIS

WHEN (GREATER-THAN (HEMATURIA (SEVERITY)) (PROTEINURIA (SEVERITY))))

where (<main-concept> (<property-name>)) is interpreted as the value of that particular property for that main-concept in the present patient. If the WHEN condition is satisfied, the comparative differential diagnosis is asserted. I haven't yet figured out how to combine this assertion with the other scores of competing hypotheses. One possible place for its use is when all findings have been entered and a diagnosis must be made. If more than one adequate hypothesis exists (see below for a complete discussion of adequate hypotheses), a global differential diagnosis may be the only criterion for choosing between them. In this particular case (and perhaps in general), however, a final decision is never made on the general level of GLOMERULITIS or G-U-TRACT-BLEEDING, but rather a choice is made between more specific examples of the categories, like FGN, AGN, PCKD, PYELONEPHRITIS etc., in which other factors are more important in making the distinction. This type of global information is certainly important for <u>explanatory</u> purposes, no matter what its use may be in processing. Dr. Kassirer has often used statements which were essentially English equivalents of the assertion above in explaining the relationship between HEMATURIA and PROTEINURIA to me.

6.7 Chore 3: Examining CHOICE-SETS

A lot of information is inherent in the designation of a CHOICE-SET and the global assembly phase may take advantage of it. If a CHOICE-SET is labelled EXHAUSTIVE and its category has been accepted, the global assembly process checks to see if all but one of the CHOICE-SET members have been rejected; if so, the remaining possibility should be accepted. For example, suppose we are sure a patient has a urinary-tract-infection (UTI), but his or her urine does not contain any bacteria. Since BACTERURIA (bacteria in the urine) is a NECESSARY EXPECTATION in BACTERIAL-UTI, this possible member of the CHOICE-SET is ruled out. However, the CHOICE-SET is labelled EXHAUSTIVE and contains only one other member: FUNGAL-UTI. The global assembly process spots this configuration, whose template is shown in Diagram 6-16 and puts the remaining CHOICE-SET member on the ACCEPTED-LIST. This is referred to in common English as "reasoning by process of elimination." Another example: HYPERTENSION CHRONIC has been eliminated as the explanation of HYPERTENSION by the absence of all of the following: RETINOPATHY HYPERTENSIVE (pathology of the retina due to high blood pressure), LVH (enlargement of the left ventricle of the heart) and ANTIHYPERTENSIVE-DRUGS (STATUS GIVEN). Since HYPERTENSION ACUTE is the only remaining member of the CHOICE-SET, it may be accepted as the explanation.

Notice that in EXHAUSTIVE CHOICE-SETS, the acceptance of the

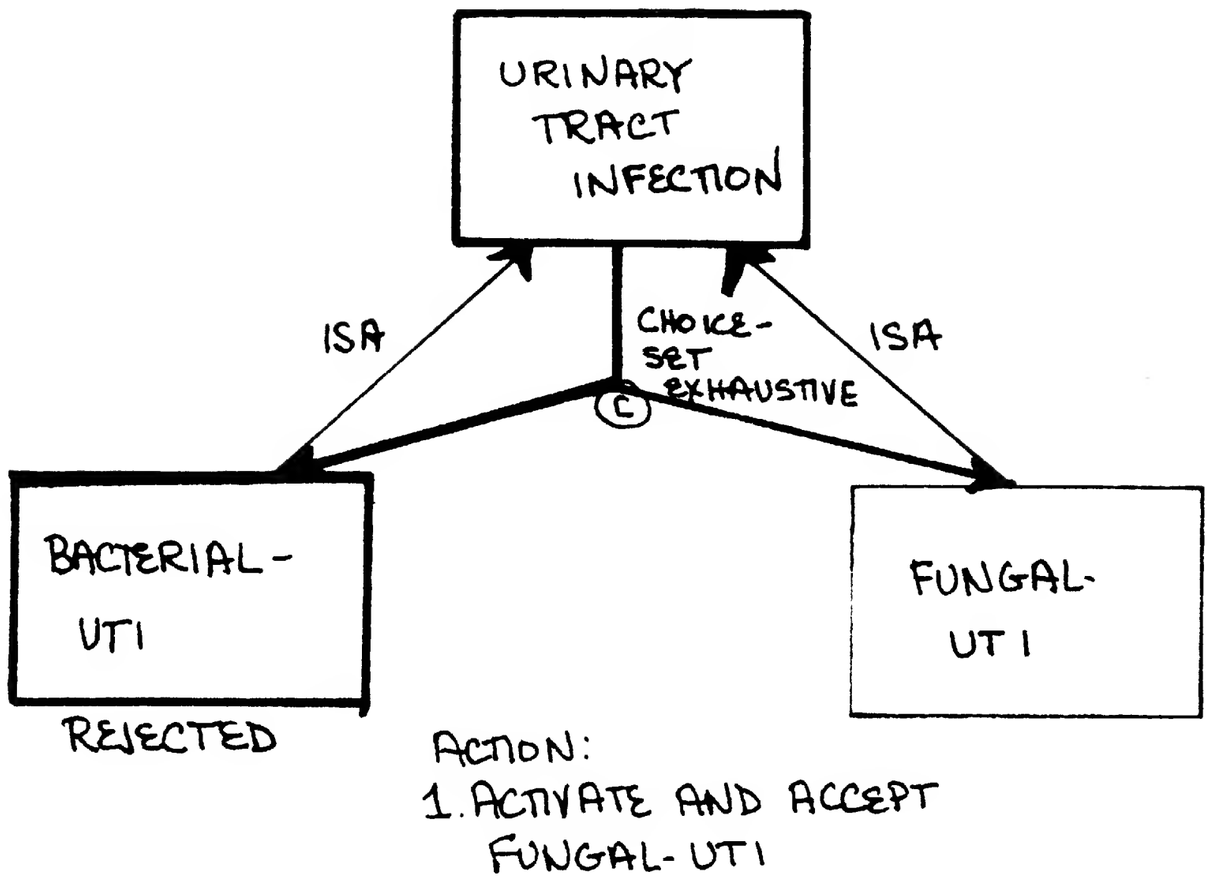


Diagram 6-16: CHOICE-SET INVESTIGATION

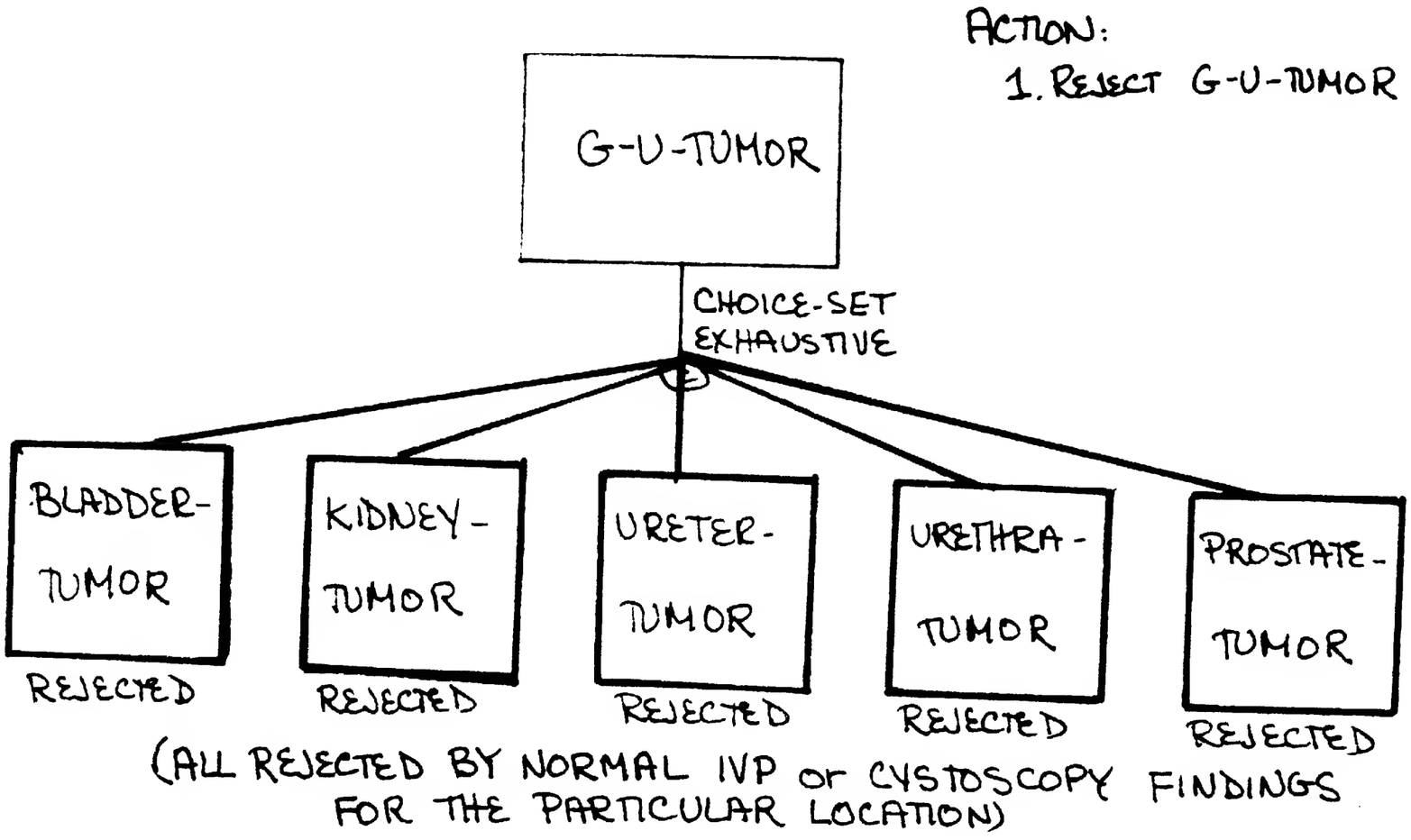
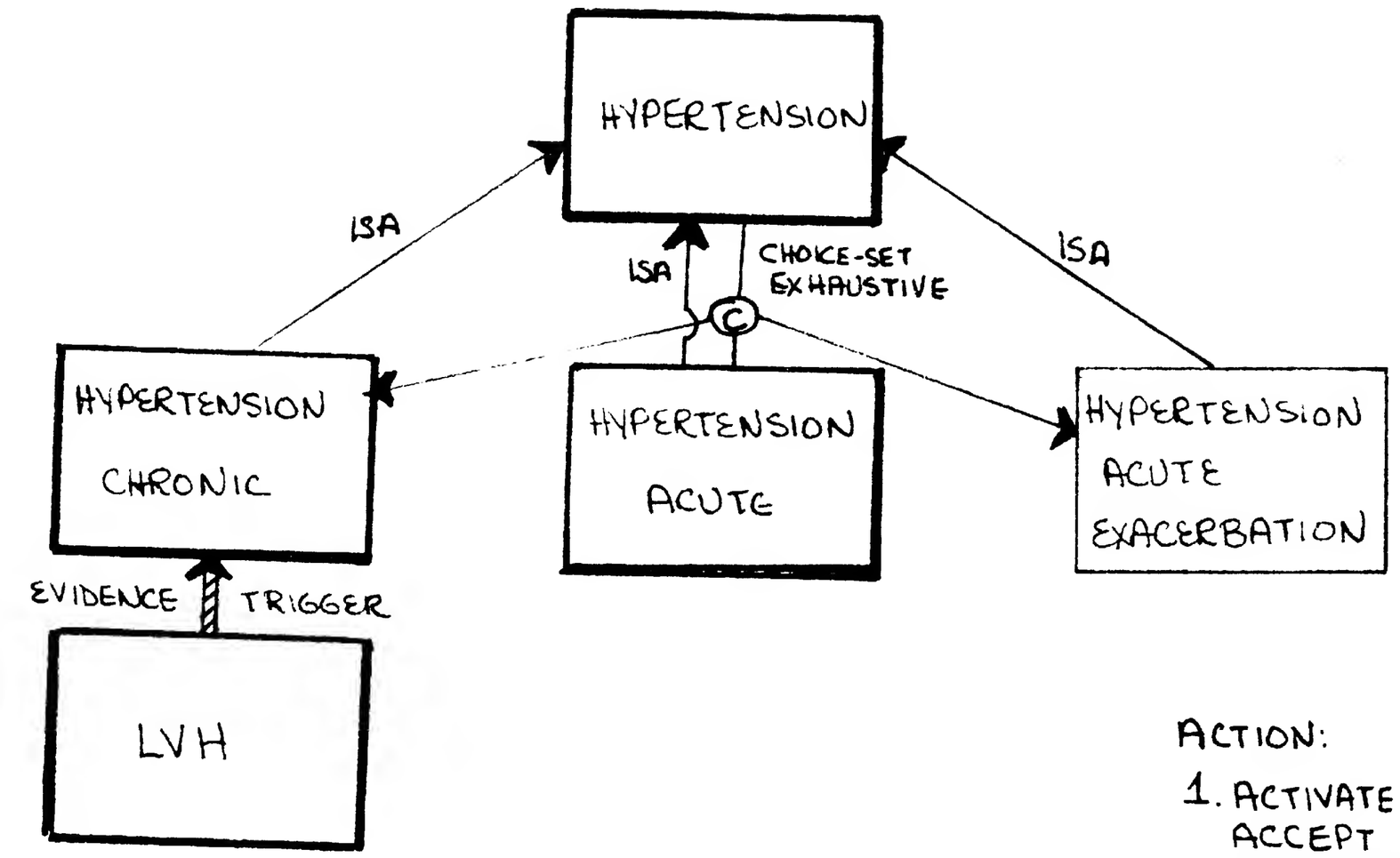


Diagram 6-17: CHOICE-SET REJECTION

category is dependent on the acceptance of one of the CHOICE-SET members. The global assembly process is also on the look-out for configurations like that in Diagram 6-17, in which a category is active but all its examples rejected. In this situation, it rejects the category as well.

I have stated above that CHOICE-SETS are mutually exclusive, meaning that the presence of one member of a CHOICE-SET rules out the presence of all the others. HYPERTENSION ACUTE and HYPERTENSION CHRONIC have been the prime examples. Clearly, if they really cannot co-exist, there should be no coherent hypothesis which contains both of them and a mechanism could easily be provided which checks this restriction every time a coherent hypothesis is formed by this processing stage. In this particular case, however, I lied - evidence of both acute and chronic hypertension may coexist and the two elementary hypotheses should be subsumed in one hypothesis postulating ACUTE EXACERBATION of CHRONIC HYPERTENSION. A patient with both LVH (see explanation a few paragraphs above) and an unusual rise in blood pressure over a short time is suffering from such an exacerbation. We might handle this phenomenon with a template illustrated at the top of Diagram 6-17a. Even though this structure is specific to HYPERTENSION, it is clearly an instantiation of a more general template structure shown at the bottom of the figure.



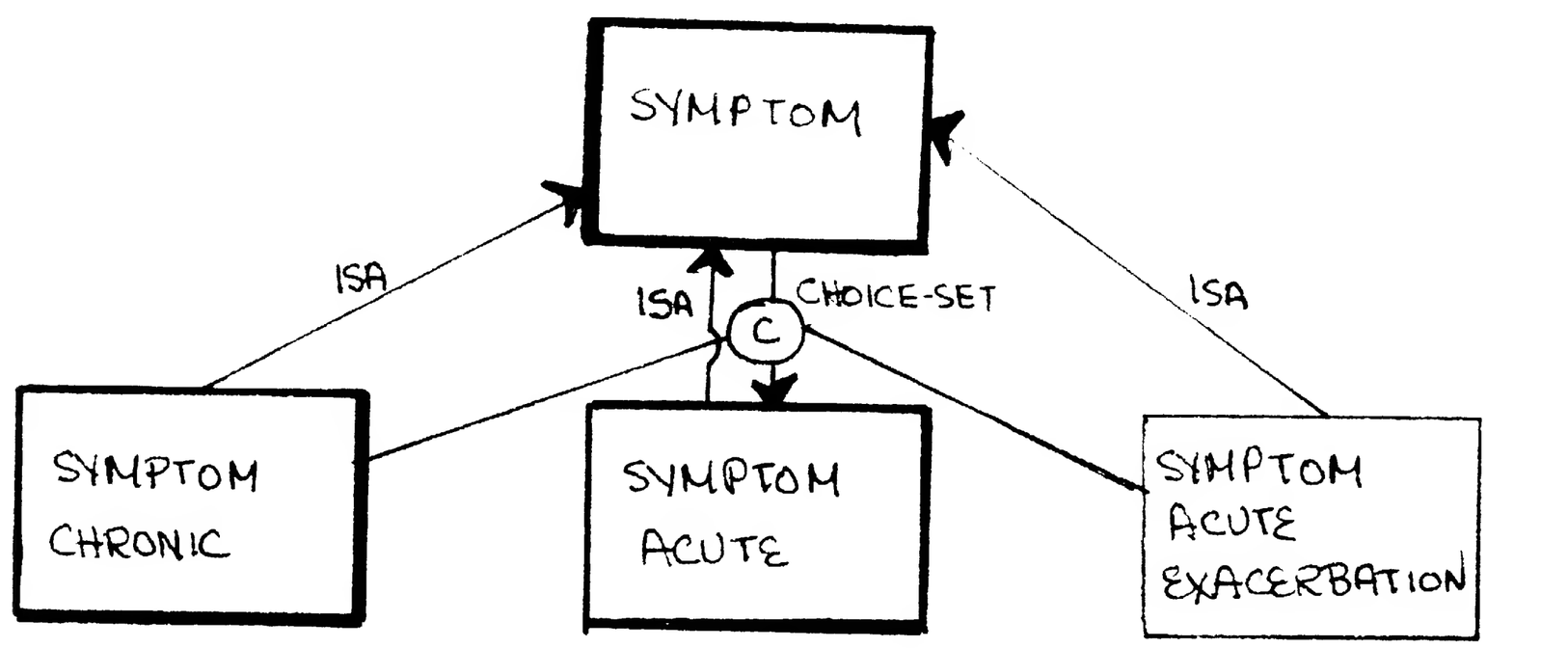


Diagram 6-17a: TEMPLATES FOR ACUTE EXACERBATION

6.8 Chore 4: Forming Adequate Hypotheses

6.8.1 Static Description of Adequate Hypotheses

I will describe the end-product of this chore in some detail, but will only make speculations about the process which might accomplish this goal, as it appears to me to be the most difficult task a diagnostic system must perform.

An adequate hypothesis is the final goal of a diagnostic procedure; the diagnosis a doctor gives at the end of a session should be an adequate one. The primary characteristic of an adequate hypothesis is that is accounts for all the abnormalities noted, while maintaining as much simplicity as possible. An adequate hypothesis consists of several independent parts, each of which is a coherent hypothesis. Each component must also be an ULTIMATE-ETIOLOGY or, in the case of more complex coherent hypotheses, it must contain some ULTIMATE-ETIOLOGY. In addition, all accepted elementary hypotheses must be subsumed in the final diagnosis, either by themselves, or as part of a larger coherent hypothesis. This is an obvious stipulation; if a doctor is sure a patient has a particular disease, it had better be part of the final diagnosis. For example, the following, taken directly from the protocol, is an adequate hypothesis:

```
LGN
        (DURATION (YEARS 10))
HYPERTENSION ESSENTIAL
        (DURATION (YEARS 5))
FAMILY-HISTORY NEPHRITIS
```

Notice that the second component of this hypothesis is HYPERTENSION ESSENTIAL, not HYPERTENSION CHRONIC; this is because only HYPERTENSION ESSENTIAL is marked as an ULTIMATE-ETIOLOGY. HYPERTENSION CHRONIC is a symptom, not an explanation, while HYPERTENSION ESSENTIAL is an explanation (actually the admission that no other explanation has been found!) All FAMILY-HISTORY and FACT findings are also considered ULTIMATE-ETIOLOGIES, so they may be included directly as a component of an adequate hypothesis.

An adequate hypothesis can in many cases be considered disease-centered, since there is often a central component which accounts for most of the symptoms and is most serious. This designation corresponds to some intuition on the part of doctors as to what the most important malfunction is and is the first disease they would mention when asked the question, "What's wrong with me, doc?" Referring to the protocol again, we note that at one point (after FINDING10 - ANTIHYPERTENSIVE-DRUGS (STATUS TAKEN)) there were two adequate hypotheses which consisted of more than one independent component - I dubbed them the LGN-centered and FGN-centered hypotheses. When the most important component of an adequate hypothesis is a complex coherent hypothesis, it may be less clear what disease to choose as the central etiology. In EVIDENCE-chained hypotheses, the choice is clearly the center as defined in Section 6.5.2 above, but in the case of CAUSE, COMPLICATION-OF and DEVELOPS-INTO connected hypotheses, the choice seems to depend on the

specific example. CLOTTING-DISORDER is a COMPLICATION-OF PREGNANCY but is more central; NEPHROTIC-SYNDROME, on the other hand, can be a complication of any GLOMERULITIS, but the particular GLOMERULITIS is clearly the center of the diagnosis. This issue may seem a bit peripheral to the problem of actual diagnosis, but I want to point out its relevance to the task of generating intelligent output for doctors from a system such as this. A diagnosis which starts out, "The patient has a hangnail and essential hypertension and by the way also has just had a severe heart attack" is obviously not acceptable.

Adequate hypotheses are ranked on two independent scales, the interactions between which I have not figured out. One measure of goodness is the number of independent hypotheses in the structure; an adequate hypothesis with fewer components is to be preferred over one with more, as it has related more of the patient's findings to each other. In the protocol, LGN DEVELOPS-INTO CGN was a better adequate hypothesis than LGN and HYPERTENSION ESSENTIAL in accounting for HYPERTENSION CHRONIC. A second scale has to do with the scores of the component parts: the comparison should probably be made between the lowest-scored component of each adequate hypothesis. Another possible comparative measure would be the average of the components' scores. The importance of this ranking could be seen in the protocol when the entire LGN/CGN centered hypothesis was dismissed because of a violated expectation in CGN which drastically lowered its score; a normal BUN level was unexpected, as CHRONIC-RENAL-FAILURE is a NECESSARY

EXPECTATION in CGN. The final diagnosis will be that adequate hypothesis which ranks highest on some combination of these scales, given some combining algorithm which I have not investigated. Sometimes in fact a doctor cannot make a choice: witness the Chapter 2 protocol.

6.8.2 Remarks on a Process to Build Adequate Hypotheses

Essentially, a process which bulds adequate hypotheses must partition the symptoms into possibly non-disjoint subsets and account for each subset with some coherent hypothesis. Clearly, such a procedure should not consider all possible partitions of the findings in regard to all the active hypotheses. What are some of the principles doctors might use to reduce the complexity of the problem?

First of all, there is the trend toward inertia. Once a doctor has considered a hypothesis for a while and has, in a sense, invested time and effort in it, he or she is reluctant to give it up. Thus, the tendency is to add new symptoms on as independent entities or attribute them to the previously-considered hypotheses, even it they are rare findings in that disease and thus may lower its score.

A finding can be added as an independent component of an adequate hypothesis under only a few conditions: FACTS, like PREGNANCY, can be added with no trouble; FAMILY-HISTORIES can as well, although an adequate hypothesis which ties one in to a patient's

present condition will be more highly ranked than one which doesn't. Other findings can be added easily only if there is some cause for them which is very common - that is, whose a priori probability is very high. In the protocol, for example, HYPERTENSION ESSENTIAL was a reasonable addition to the LGN-centered hypothesis because its a priori probability in a 31-year-old woman is high. Similarly, a bloody nose or a headache might be independent components of an adequate hypothesis because they occur commonly.

Another trick in reducing the complexity of forming adequate hypotheses is to dispose of a finding as soon as it appears by attributing it to an obvious cause rather than considering it in relation to several possible elementary hypotheses. This has been explained as the first processing step in Chapter 3. To repeat the example cited there, a doctor should attribute HEMATURIA in an accident victim to trauma and not think about all the other possible causes for the abnormality.

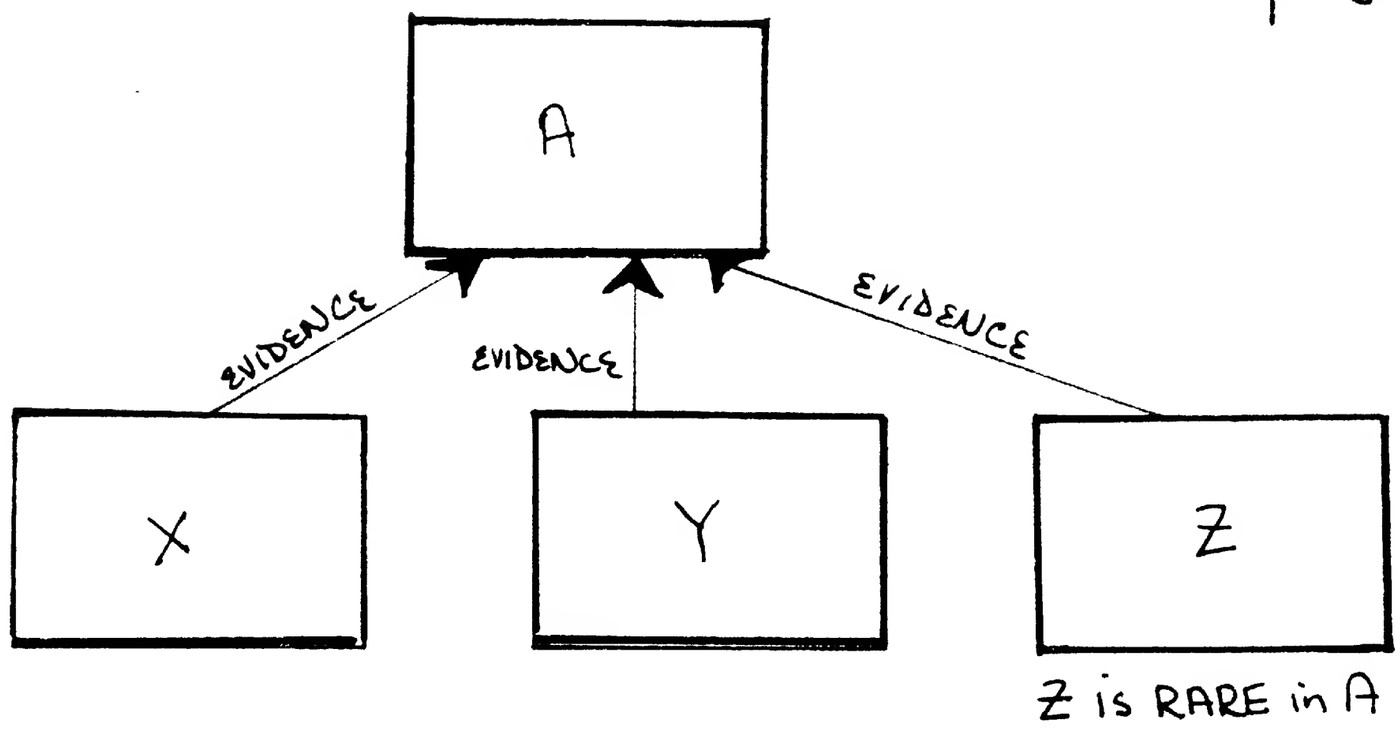
Even with these techniques, a process which forms adequate hypotheses will have to be able to do some fiddling of the following sort: Suppose we have hypothesized disease A to account for symptoms X, Y, and Z. Z is a rare finding in A, so considering it relevant to A will lower that hypothesis' score. Disease C accounts for symptoms Q, R and Z; Z is a common finding in C. Putting A and C together in an adequate hypothesis which accounts for all the symptoms - Q,R,X,Y, and Z - allows us to attribute Z to C and not to A; thus A's score should

be higher. Diagram 6-18 is a schematic representation of this situation; notice that it too is a template, but the action is to combine the two hypotheses into an adequate hypothesis which accounts for all the symptoms more satisfactorily. Clearly, then, forming adequate hypotheses can affect the individual hypotheses' scores, so the combining problem may become complex.

The process of forming adequate hypotheses assumes more importance as the diagnosis proceeds; at the beginning, a doctor probably does not always concern him or herself with concocting the best explanation for all the symptoms. It is more important for the first few symptoms to trigger new hypotheses (but not too many!) which organize the findings into chunks and provide the doctor some idea of what the patient's problem could be. Later, when more symptoms have accumulated, the doctor must "finetune" his or her hypotheses and at this later stage the formation of adequate hypotheses assumes greater importance; this is one aspect of the previously-mentioned switch from symptom-centered to disease-centered processing (see Section 3.4.7).

6.9 Summary

This chapter dealt with the final step in the processing cycle which follows the addition of each finding: global assembly. After discussing the most local type of coherent hypothesis - one which includes several time-instantiations of the same disease - more global



ACTION:
1. FORM ADEQUATE HYPOTHESIS SHOWN BELOW

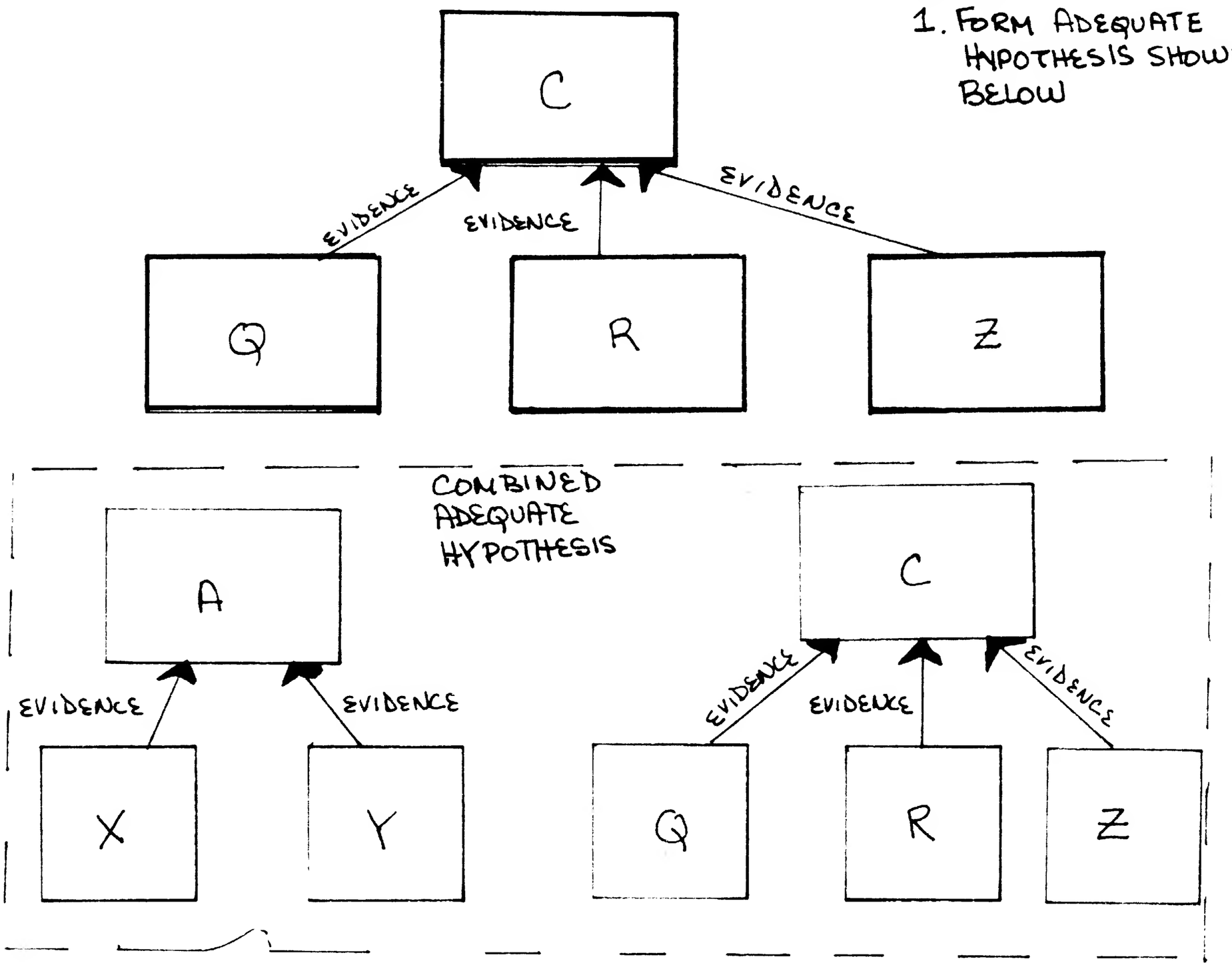


Diagram 6-18 A TEMPLATE FOR FORMING ADEQUATE HYPOTHESES

structures were considered.

The designation of pathological states like SODIUM-RETENTION which can occur in many diseases as separate elementary hypotheses was seen as a mechanism for aiding a doctor's memory of disease patterns and allowing a single faulty mechanism to be hypothesized outside the context of a specific disease. This separation of symptom from disease by intermediate elementary hypotheses, however, necessitated special mechanisms to re-unify several elementary hypotheses relating to one disease into a coherent hypothesis.

The coherent hypothesis templates were of four types: ISA-connected, EVIDENCE-chained, CAUSE, COMPLICATION-OF or DEVELOPS-INTO connected and EXCUSE-connected. If any of the templates described fits the patient data-structure, the specified action is taken; this always involves joining the matched elementary hypotheses into a more complex coherent hypothesis and may in addition require the activation of new elementary hypotheses. Forming coherent hypotheses is the first chore of the global assembly stage. A comparison between the use of these coherent hypothesis mechanisms and multiple triggers revealed differences in terms of uniformity/specificity and also pinpointed multiple triggers as yet another example of local compilation of global knowledge.

The second and third chores dealt with global differential diagnosis and CHOICE-SETS, respectively; both these are clearly global processes because they deal with more than one elementary hypothesis

at a time.

The fourth and final chore of the global assembly stage was the formation of adequate hypotheses, the end products of a diagnositc session and the set from which the final diagnosis is chosen. While I gave a fairly detailed description of the properties of an adequate hypothesis - most important, that it accounts for all the data - the procedure for carrying out their formation was only sparsely specified and remains a problem for further research. It seems, though, that templates similar to those used in recognizing and forming coherent hypotheses can be used in the process of putting together adequate hypotheses.

Chapter 7 - Reflections, Retractions and Reveries

This chapter, in the tradition of all final thesis chapters, looks back on the previous six chapters and comments on their significance. It examines the theory developed here in relation to recent developments in A.I., points out some conceptual difficulties with some of the theory's approaches and speculates on possible future developments. The reflections tackle the relationships between "frame" theory and this theory of medical diagnosis. The retractions primarily concern the local evaluation algorithm and the notions of EVIDENCE and EXPECTATION. The reveries consider some possible implications for the process of gaining expertise or learning.

7.1 Reflections

A frame has recently been described by Minsky as "a remembered framework to be adapted to fit reality by changing details as necessary." <Minsky 74> A frame thus represents an abstraction from reality - it is a structure which does not represent a single entity, but a prototype into which real objects can be fit. In daily life we constantly make the correspondences between a frame and some portion of our experience; in vision, this constitutes recognition of the object: in language, understanding the phrase or sentence. Since each

frame contains general information about the class of entities it represents, deciding a frame applies to a situation allows one to use all that general information, even though much of it may not be immediately derivable from the data itself. For example, if we decide a particular object fits the car frame, we believe it has four wheels, even though only two of them may be visible. If we further decide it fits the Cadillac frame because of its size and lines, many other implications follow: it has a V-8 engine, carpeting in the interior etc.

Typically, instantiating a frame by deciding that the current situation fits it involves filling in certain slots in the frame with more specific descriptions called fillers. Minsky's prime example of this is a room frame. It contains four slots corresponding to walls and one each for the ceiling and floor. When three of the four wall slots are filled with specific descriptions of the walls a person sees, he or she also knows a fourth wall exists behind the field of view, which will be visible if he or she turns around. The act of turning 180 degrees is mirrored by a transformation to a new frame which shares slots with the original one; the new frame asserts that a different wall is invisible and that the right and left walls are appropriately interchanged. Minsky has called a collection of such related frames which share slots (also called terminals by Minsky) a Frame-System.

Besides being connected by links which correspond to changing

viewpoints, frames are organized hierarchically, with the most general structures appearing near the top and the most specific at the bottom. For example, a general room frame contains very little information about the specific descriptions of the walls or contents of the room. A bedroom frame, however, is found below the general room frame in the hierarchy (it ISA room) and contains more specific information about the characteristics of the room: it must have a bed, etc. The process of making a frame more specific has been called (by Newell <Newell 73> and Winograd <Winograd 74>, among others) _further specification_.

The process of further specification goes on both in defining frames as examples of others (the room/bedroom example) and in instantiating frames by fitting them to the real world (relating the bedroom frame to a real bedroom by filling in the slots.) Notice that in this way, frames provide methods for organizing and structuring both knowledge in long-term-memory (LTM) and incoming information, which is traditionally thought of as residing in short-term-memory (STM). The organization of knowledge into uninstantiated frames makes of LTM more than a mass of disconnected assertions. The fitting of "reality" to frames imposes a structure which is necessary to "understand," as well as facilitating memory of the data which, without organization, would soon fill up the small number of available places in STM.

Another crucial aspect of frame theory is the designation of _default values_ for various slots, values which are assumed to be

representative of "reality" unless contradicted by the data. I will take an example from the domain of language understanding, as that is another area which has recently been explored using the "frame" metaphor. The frame for the verb (or action) "drive" contains a default value of "car" for the "vehicle" slot; it can be overriden, however, by explicit mention of another vehicle, as in

She drove to San Francisco in an orange truck.

One final mechanism proposed originally in the context of frames is a "trigger" - a concept which suggests the relevance of a particular frame in structuring the current situation. The trigger -> frame mapping process may be as simple as a verb triggering the verb/action frame whose name it shares ("drive" -> drive frame) or may be more complicated, as one of the eventual slots-fillers may suggest a frame ("cake" -> birthday party frame). Since each slot can also be thought of as a frame, the activation of a frame requires the activation of all frames corresponding to its slots, as birthday party -> trigger "present" frame. A discrepancy between expectations inherent in the selection of a given frame and actual details of the data may also trigger another frame which might fit better.

How do these general aspects of a theory for representing knowledge fit in with the theory developed here for structuring and using medical knowledge? On this general level, the similarities are clear, partly because the medical theory has drawn heavily from the frame theory developed earlier. Elementary hypotheses clearly

correspond to frames; they organize data into more structured chunks, and provide a basis for expectation. When an elementary hypothesis has been activated, its other relevant symptoms are expected to be present. Elementary hypotheses may be triggered by an appropriate combination of relevant symptoms, causes, complications etc. or by explicit mention in the case of an expectation/fact discrepancy noted in another elementary hypothesis. An elementary hypothesis's slots are its relevant symptoms.

I have actually used the phrase "further specification" above in Chapter 3 in referring to the relationship between a slot-description and its potential filler; a further specification is one which matches the slot-specification in all possible ways and, in addition, contains more information about other properties such as severity, location etc. The notion of further specification also comes out clearly in the CHOICE-SET designations - every member of a CHOICE-SET is a further specification of its category. Sometimes the name of the slot being filled is obvious: going from G-U-TUMOR to KIDNEY-TUMOR involves making the filler of the "location" slot more precise. Instantiating an elementary hypothesis in the course of a particular diagnosis also implicitly involves filling a slot called "patient." I have not emphasized this distinction between the uninstantiated knowledge network which is general with respect to patient and the instantiations created when a particular person is being diagnosed. It is clear, however, that we may want to consider

several different occurrences of the disease in different people during one diagnostic session, especially, for example, in the case of a hereditary disease. These occurrences should be considered separate instantiations of the disease-hypothesis, each with a different filler for the "patient" slot. I have mentioned instantiation most explicitly in regard to time-instantiations; it should be apparent that they are just one example of a more general concept which originally comes from frame theory.

Another similarity between the two theories is the notion of default and the possibility of overriding a default. In frames, an example of a default is a slot-filler which can be overriden by real-world data. In the medical diagnosis theory, however, a default assumption is that a symptom-cluster can be evaluated independently of the disease in which it occurs. (see Section 5.3 on the "X" phenomenon.) and an OVERRIDE is an explicit rejection of that convention by mention of a symptom/disease interaction which violates that default procedure.

But even with these very general similarities, there appear to be some conceptual differences between frames in a vision or language system and elementary hypotheses in a medical diagnosis system. For one thing, the emphasis is different. In diagnosis, we are mostly concerned with deciding what hypothesis fits the data, what malady is afflicting the patient; triggers are helpful in keeping the number of active hypotheses to a minimum, but are seldom diagnostic. In

contrast to this, language frames, for example, are often chosen by a single word (or word-class) and the emphasis is on filling the slots correctly with other entities described in the sentence or larger linguistic context. Sometimes those fillers can help to disambiguate several senses of a word, as in:

She plays the tuba for the high school band.

She plays football for the high school team.

but the process seems significantly different from those involved in medical diagnosis. At the grossest level, probabilities and indefinite decisions which play a central role in medical diagnosis are not used in text (written language) understanding programs. However, a new emphasis on speech (oral language) understanding has brought probabilistic methods back into consideration because the acoustic signal is so hard to decipher and many hypotheses must be entertained at once.

A second difference has to do with the number of fillers a slot will accept. The slot-specifications in, for example, ACUTE-RENAL-FAILURE are highly specified - (SERUM-CREATININE (RANGE RISING)), (BUN (RANGE RISING)), (URINE-VOLUME (RANGE LOW)) etc. In a frame such as that for a room or a verb, however, the restrictions on slot-fillers may be simply expressed as markers like VEHICLE or WALL-DESCRIPTION. The patient and location slots in disease frames are more similar to language slots in the range of fillers they admit. When a slot may be filled by many entities, the structure

frame-with-slot-filled often attains independent conceptual significance - as "shoot with an arrow." ("shoot" with the instrument slot filled with "arrow"), "Mr. Hypochondriac's sore throat" ("sore-throat" with the patient slot filled) or room with a glass wall. In language, such partially-instantiated structures are often immortalized in a word, like the verb "bus" ("drive" with the instrument slot filled with "bus"). It is hard to fit the symptom-slots of elementary hypotheses into this view of the role of slots and fillers. I think, in fact, that they are significantly different, especially in the probabilistic role symptoms play in choosing the correct frame or elementary hypothesis to apply to a particular patient.

Frames for recognizing visual objects and scenes will probably turn out to share more with medical frames. They are often triggered by a feature or attribute of the situation, like a horn triggering BULL; these triggers are far from "diagnostic," though, since the horn could belong to a unicorn or a rhinoceros. A lot of the recognition process involves discovering whether properties are PRESENT or ABSENT, much as in medical diagnosis. The emphasis, unlike that in text understanding, is really on finding the right frame or group of frames to account for all the observed phenomena. In some types of visual pattern recognition, in fact, probabilistic methods similar to those explained here have been used. Fahlman <Fahlman 73> is currently investigating frames for the visual recognition of animals and their

relationship to other frame-like systems, including that presented here.

7.2 Retractions

There are a few areas of the theory described here which, after more thought, I have decided are incorrect or unintuitive. For people who are planning to on in the investigation of medical diagnosis, this section is most important, for progress in any field involves not repeating others' mistakes.

I originally included the concepts EVIDENCE and EXPECTATION because of my feeling, explained in detail in Chapter 4, that there is a significant difference between what I called disease-centered information and symptom-centered information. While that difference certainly exists, its translation into EVIDENCE values which always add to the score of an elementary hypothesis when a symptom is present and EXPECTATION values which indicate the amount to subtract from a hypothesis when an expected symptom is absent has confused two separate issues. One issue is whether or not a disease hypothesis can account for a symptom; I have called those symptoms a disease can account for the relevant symptoms in its slice. Just because a disease can cause a symptom, however, does not mean that its presence is positive evidence for that disease's existence. What it boils down to is: how do we express the fact that a symptom is RARE but possible

in a disease; for example a bloody nose is rarely caused by the flu, but the possibility exists. This type of fact is precisely the "raw data" with which a doctor starts out - the probability of symptom given disease - and it seems to be more useful than I indicated in Chapter 4. Within the outline of the theory I have developed, a solution to this problem is not too difficult - all symptoms included in a disease's slice can be accounted for by it, unless specially marked as PRECLUDING it or as suggesting another hypothesis via a differential diagnosis. The presence of a symptom, however, may add less evidence than its absence - and may even subtract from the hypothesis' score. The normalizing number by which a raw score is divided is still the highest score an elementary hypothesis could have, considering just the symptoms about which we have information - but the highest score may reflect the absence of some symptoms and the presence of others, rather than the presence of all of them.

I will not further develop the mechanisms of this slight change, however, because the whole scoring and local evaluation procedure seems to me to be slightly misguided. Even with the simplification I made to four strengths of EVIDENCE and EXPECTATION, the scores and the combining algorithms became complex. It seems unlikely that doctors really use such complicated arithmetic operations in evaluating the possibility of a disease's presence.

There are technical problems with the approach to scoring developed here as well. I ran into a problem in working through the

protocol, trying to simulate Dr. Kassirer's rejection of KIDNEY-STONES upon hearing the finding FLANK-PAIN ABSENT. The hitch was that, no matter how much the absence of flank pain subtracted from KIDNEY-STONE'S score (up to a maximum of 1), it could not counteract the positive contribution of HEMATURIA and PROTEINURIA. Short of proclaiming FLANK-PAIN a NECESSARY EXPECTATION of KIDNEY-STONE (which it isn't, since there are cases of KIDNEY-STONE which occur without pain), there was little I could do to make the numbers come out right. What I wanted, obviously, was a way to consider how serious a discrepancy the lack of pain was, independent of how much other evidence for KIDNEY-STONE there was.

In addition, playing with numbers necessitates keeping in mind the relationships between different somewhat arbitrarily assigned weights. For example, in trying to decide how much evidence HEMATURIA MICROSCOPIC and HEMATURIA GROSS contribute to GLOMERULITIS, I had to be aware of the ratio between them and their relationships to the (negative) contribution of HEMATURIA ABSENT, as well as all the other symptoms of GLOMERULITIS.

It's clear that a simpler, even less combinatorial theory is desirable and though I have not worked one out in any detail, its outlines follow.

Symptom-centered information is especially important in two situations: for triggering and for accepting hypotheses. A hypothesis is generally triggered by a symptom if it is an especially frequent

cause for it or is an especially common disease. (Recall the more mathematically based discussion in Chapter 4.) Determining triggers requires global knowledge, as it requires comparing several possible diseases to choose the most common; thus, the designation of a symptom as a trigger for a disease is a local compilation of this global knowledge. In addition, finding a symptom or combination of symptoms which are SUFFICIENT EVIDENCE for a disease is important because it enables a doctor to accept a hypothesis on the basis of a few symptoms without having to examine all other possible etiologies. So the new theory keeps these two forms of derived quantities, as they were developed above.

Central to a better theory is the concept of discrepancy, which encompasses both what I have called violated expectations and unaccounted-for symptoms. There may be a severity associated with each discrepancy. In the violated expectation case, some absent symptoms may be more common than others and their absence thus more worrisome. Certain symptoms may occur VERY-RARELY or NEVER in a disease and their presence is thus a more serious discrepancy than the presence of a symptom which is more common. All the interactions catalogued in Chapter 5 will still be operational, as well.

From this point of view, we can define the prototype of a disease as the set of symptoms which contains no discrepancies - if a symptom is more often present than absent, it will be present in the prototype; if it is rare, it will be absent. There may be several

prototypes for a disease, representing its different courses in, for example, different age categories. Evidence for a theory which concentrates on discrepancies from prototypes comes from doctor's remarks in protocols, for example: "You may have LGN, but 10 years is a long time to have gross hematuria in that disease." or "I would have expected flank pain if you really had pyelonephritis." The hope is that a diagnosis could be reached in this way without complex arithmetic operations. Symptom-centered information would be used primarily to suggest hypotheses (most noticeably, age-sex-presenting symptom combinations); disease-centered information would be more helpful in evaluating hypotheses and, eventually, in rejecting many of them because they had too many discrepancies with the data.

The trend from the earliest theories to the one sketched here moves uniformly away from complex probabilistic manipulations through a somewhat simplified scoring algorithm to a theory which tries to avoid as much arithmetic as possible. Although I am not sure of the success of the prototype/discrepancy theory suggested here, I feel it is a step in the right direction.

7.3 Reveries

What can we say about the process of a doctor's gaining the expertise allegedly modeled by the theories suggested here? What magic things happen to a doctor between his or her graduation from

medical school and emergence as a full-fledged diagnostician?

First, it seems clear that, although learning the medical facts is a large part of a doctor's training, it is not everything. Learning does not consist only of gaining the information, but also of organizing it, finding efficient and appropriate access paths to it, and devising useful procedures to process it.

The major part of the "making of a doctor" seems to be the derivation of symptom-centered information which provides the doctor the ability to respond to the mention of a symptom with a list of diseases it might indicate. At first, the process is probably not very discriminating, as if the cross-index links were just being derived, but different strengths of suggestion had not yet been differentiated. As a doctor's expertise develops, two opposing processes go on. He or she learns about more diseases and thus can respond with more possible causes to the mention of a symptom. On the other hand, this burgeoning number of possibilities will tax his or her memory unduly and make diagnosis more difficult, so the triggering algorithm must become more precise, activating only the most likely hypotheses, setting up multiple triggers etc. In our investigations, we have found that experts tend to face this problem by jumping to conclusions and later modifying their guesses by pre-compiled differential diagnosis pointers (explained in Chapter 5); another feasible approach is to keep the activated hypotheses at a general enough level that there are relatively few of them until more data is

available to dispose of some of the more specific ones and we should expect some doctors to adopt this strategy as well.

A systematic study of diagnostic styles at different levels of the development of expertise is being carried out by Peter Miller <Gorry 74> and might be expected to reveal the following developmental profile of concurrent active hypotheses in response to a presenting symptom (plus age and sex): first, only a few, fairly general disease categories, especially in a medical student with little clinical experience; then, as a cross-index begins to develop, an increase in both the number and specificity of activated hypotheses; finally, as the triggering process becomes more discriminating, a smaller, more precise list of possibilities. The investigation should also be sensitive to the possible influence of particular cases a doctor has diagnosed and treated on the development of prototypes; at some point, a theory of reasoning from particulars by analogic processes may be useful in following this aspect of a doctor's maturing diagnostic style.

7.4 Summary

No thesis is complete without the suggestion that further research into its subject matter should produce even more valuable and interesting results. In this case, it should be painfully obvious that much more work is needed to produce even a barely adequate theory

of medical diagnosis. In particular, the [illegible] pointed out in this thesis - the search for [illegible] can limit the number of [illegible] should prove to be a valuable guide for future [illegible]

Bibliography

<Amarel 73>

Amarel, Saul et al, "Second Annual Report of the Rutgers Special Research Resource on Computers in Biomedicine," Livingston College, Rutgers, September 1973.

<Brown 74>

Brown, Allen L., "Qualitative Knowledge, Causal Reasoning, and the Localization of Failures," WP-61, Artificial Intelligence Lab, Cambridge, March 1974.

<Card 74>

Card, Stuart K., "Encoding Knowledge about Psychological Experiments," PhD. Thesis, Department of Psychology, Carnegie-Mellon University, February 1974.

<Fahlman 73a>

Fahlman, Scott, "Does Vision Need A Special Purpose Language?" WP-50, Artificial Intelligence Lab, M.I.T., Cambridge, September, 1973.

<Fahlman 73b>

Fahlman, Scott, "A Hypothesis-Frame System for Recognition Problems," WP-57, Artificial Intelligence Lab, M.I.T., Cambridge, December, 1973.

<Feller 68>

Feller, William, An Introduction to Probability Theory and Its Applications, Vol. 1, John Wiley and Sons, Inc.: New York, 1968.

<Goldstein 74>

Goldstein, Ira, Understanding Simple Picture Programs, AI-TR-294, Artificial Intelligence Lab, M.I.T., Cambridge, June 1974.

<Gorry 67>

Gorry, G.A., A System for Computer-Aided Diagnosis, TR-44, Project MAC, M.I.T., Cambridge, September 1967.

<Gorry 73>

Gorry, G. A. et al, "Decision Analysis as the Basis for Computer-Aided Management of Acute Renal Failure," American Journal of Medicine 55, October 1973, pp. 473-484.

<Gorry 74>

Gorry, G. A. et al, Computer Laboratory for Clinical Decision-Making, research proposal to NIH, March 1974.

<Greenblatt 69>

Greenblatt, Richard, Donald Eastlake, and Stephen Crocker, "The Greenblatt Chess Program," AIM-174, Artificial Intelligence Lab, M.I.T., Cambridge, April 1969.

<Hollerbach 74>

Hollerbach, John, "Hierarchical Shape Description by Selection and Modification of Prototypes," M.S. Thesis, Department of Electrical Engineering, M.I.T., Cambridge, September 1974.

<Kuipers 74>

Kuipers, Ben, "An Hypothesis-Driven Recognition System for the Blocks World," WP-63, Artificial Intelligence Lab, M.I.T., Cambridge, March 1974.

<McDermott 74>

McDermott, Drew, "Assimilation of New Information by a Natural Language-Understanding System," AI-TR-291, Artificial Intelligence Lab, M.I.T., Cambridge, February 1974.

<Minsky 74>

"A Framework for Representing Knowledge," in The Psychology of Computer Vision, P.H. Winston, ed., McGraw-Hill: New York 1975.

<Newell 73>

Newell,Alan and J. Moore, "How Can Merlin Understand?,"
in L. Gregg (ed), Knowledge and Cognition, 1973.

<Newell and Simon 72>

Newell, Allen and Simon, Herbert, Human Problem Solving,
Prentice-Hall, Inc.: Englewood Cliffs 1972.

<Rubin 73>

Rubin, Andee, "Anti-Numerical Analysis," Artificial Intelligence Lab, M.I.T.,
Cambridge, May 1973.

<Schwartz 70>

Schwartz, William B., "Medicine and the Computer - The Promise and Problems
of Change," New England Journal of Medicine 283, December 3, 1970, pp. 1257-1264.

<Schwartz 74>

Schwartz, W.B. et al, "Decision Analysis and Clinical Judgment,"
American Journal of Medicine 55 October 1973, pp. 459-472.

<Shortliffe 74>

Shortliffe, Edward H. et al, "A Therapy Consultation Program Which Explains
the Reasoning Behind Its Advice," Stanford University School of Medicine,
Stanford, 1974.

<Silverman 73>

Silverman, Howard, "A Digitalis Therapy Advisor,"

M.S. Thesis, Department of Electrical Engineering, M.I.T., Cambridge,

December 1974.

<Strauss and Welt 63>

Strauss, Maurice and Louis Welt, (Eds.) Diseases of the Kidney,

J. and A. Churchill Ltd.: London 1963.

<Sussman 73a>

Sussman, Gerald, A Computational Model of Skill Acquisition,

AI-TR-297, Artificial Intelligence Lab, M.I.T., Cambridge, August 1973.

<Sussman 73b>

Sussman, Gerald,"Some Aspects of Medical Diagnosis,"

WP-56, Artificial Intelligence Lab, M.I.T., Cambridge, December 1973.

<Winograd 72>

Winograd, Terry, "Thoughts on Human Problem Solving,"

Artificial Intelligence Lab, M.I.T., Cambridge, October 1972.

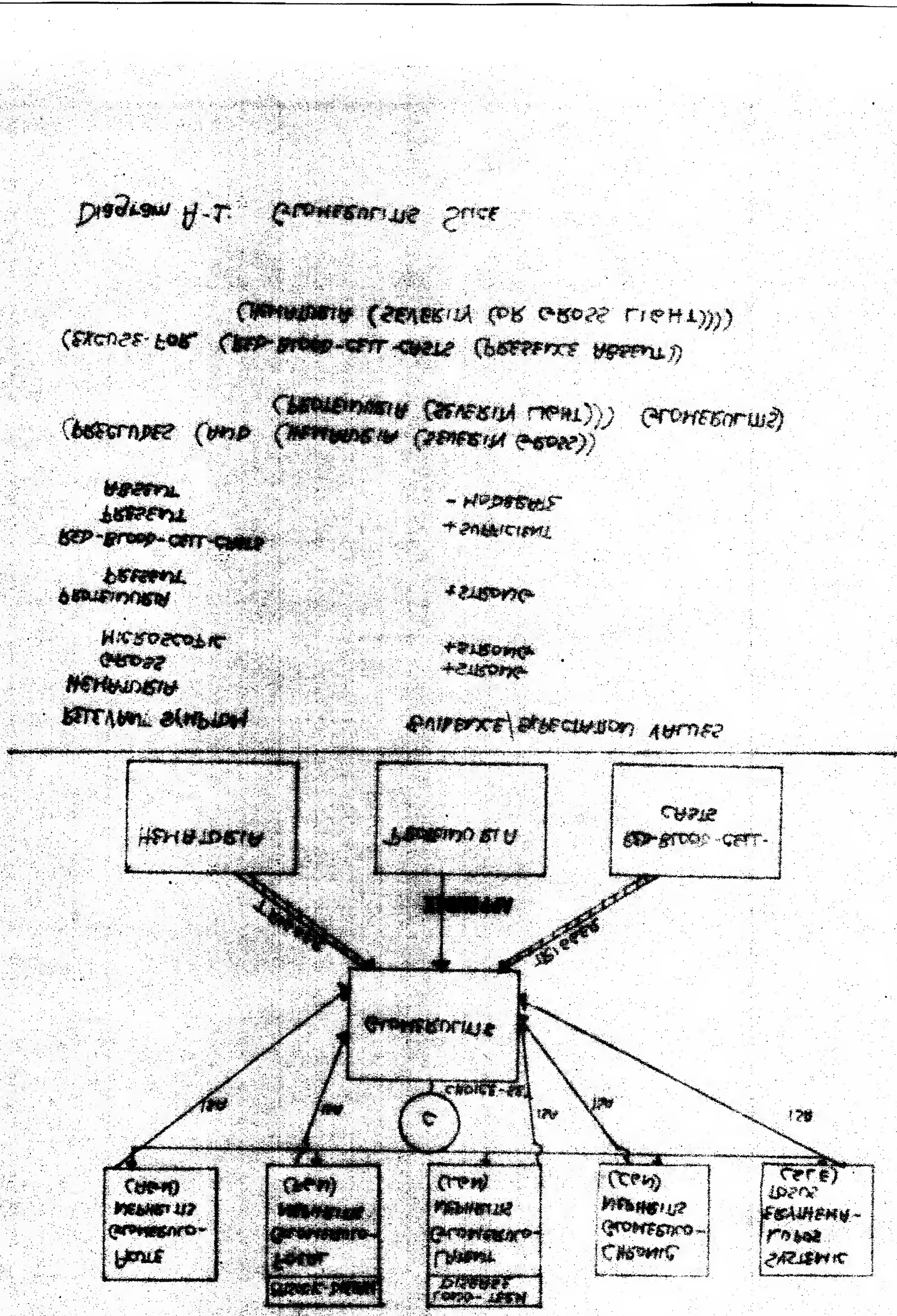
<Winograd 73>

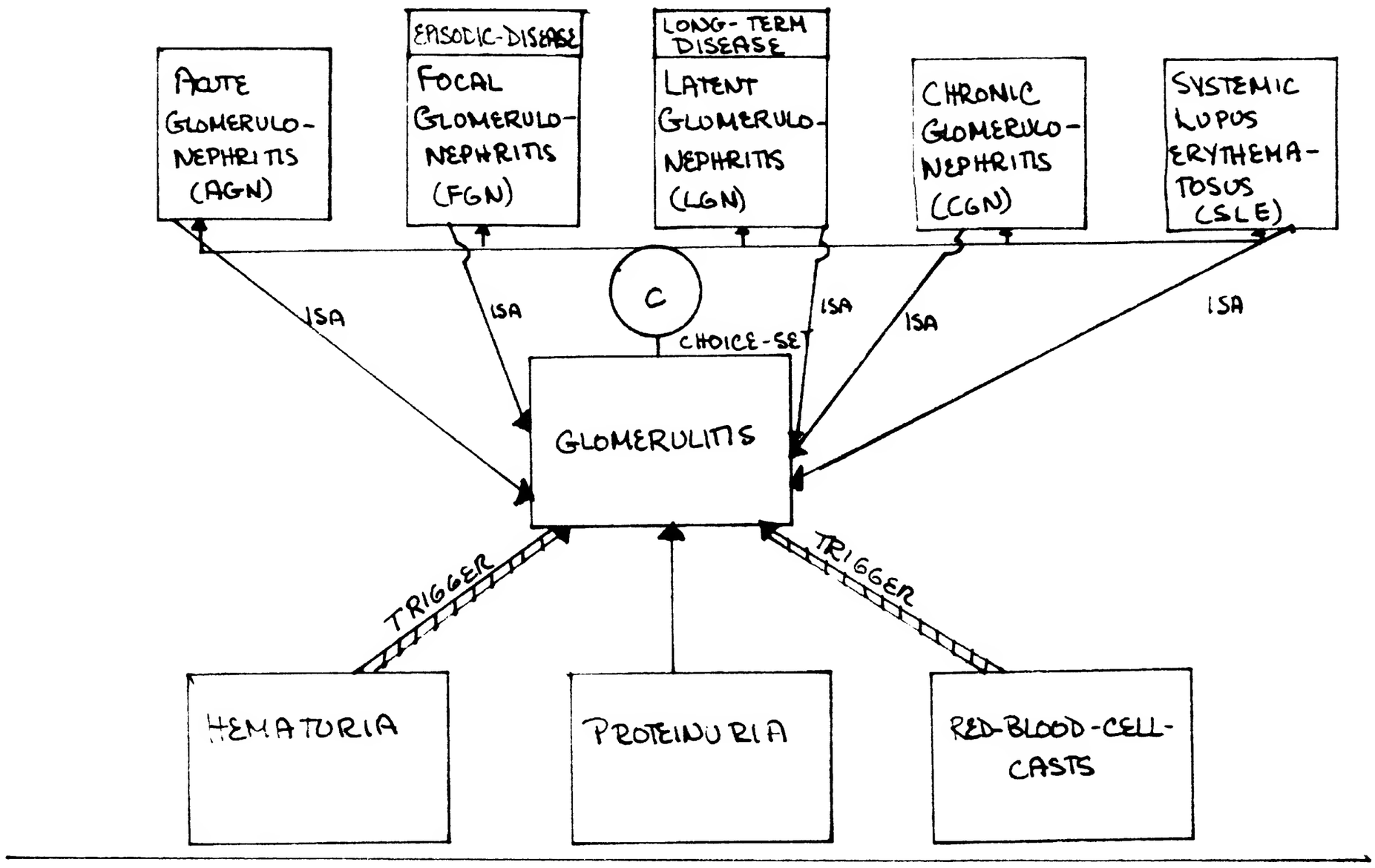
Winograd, Terry, "Breaking the Complexity Barrier - Again," In Proceedings of ACM SIGPLAN-SIGIR Interface Meeting, Gaithersburg, Maryland, November 1973.

<Winograd 74>

Winograd, Terry, "Frame Representations and the Declarative/Procedural Controversy," in Representation and Understanding: Studies in Cognitive Science, Academic Press, 1975.

Diagram H-1: Conference Call





RELEVANT SYMPTOM — EVIDENCE/EXPECTATION VALUES

HEMATURIA
- GROSS — +STRONG
- MICROSCOPIC — +STRONG

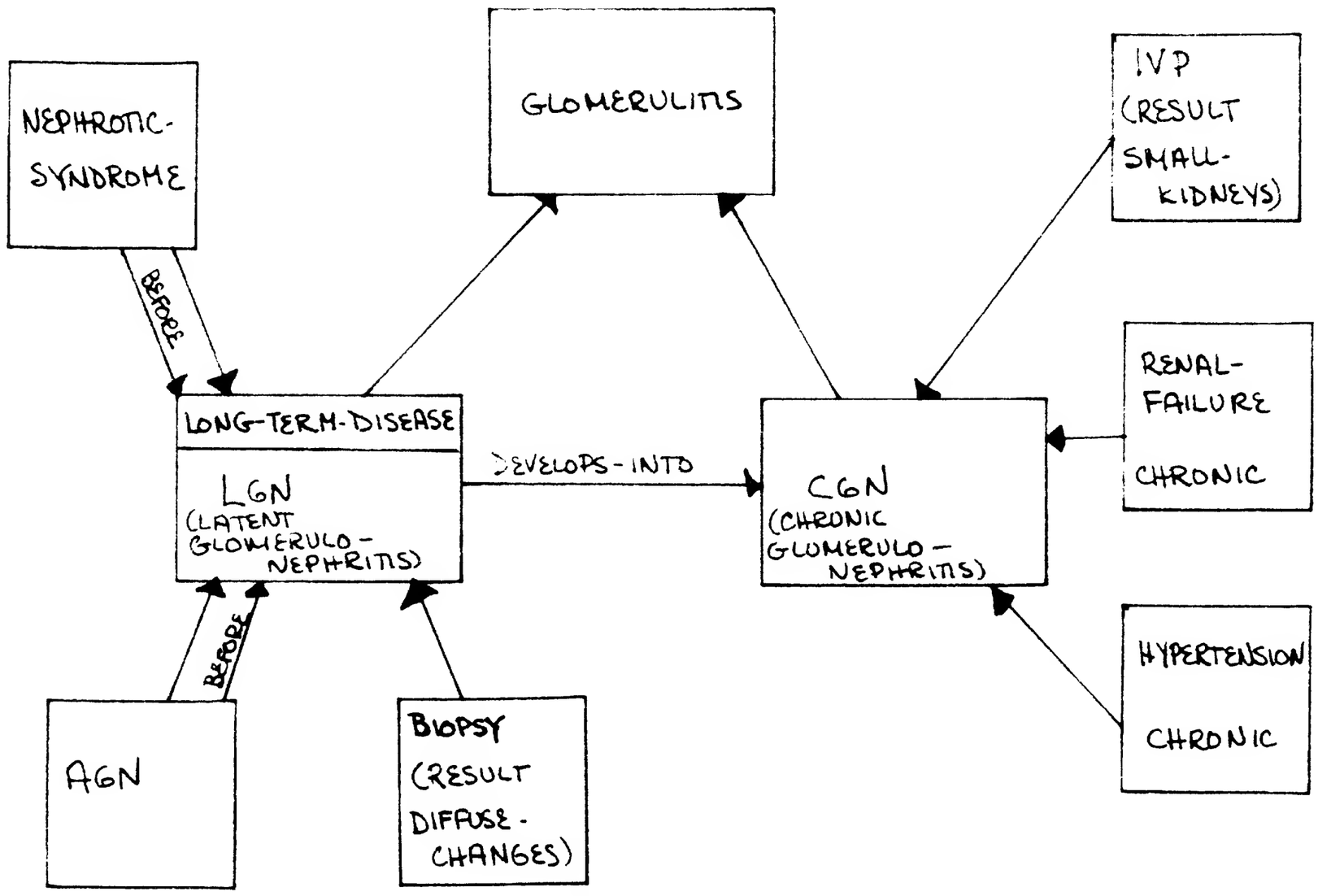
PROTEINURIA
- PRESENT — +STRONG

RED-BLOOD-CELL-CASTS
- PRESENT — +SUFFICIENT
- ABSENT — -MODERATE

(PRECLUDES (AND (HEMATURIA (SEVERITY GROSS))
(PROTEINURIA (SEVERITY LIGHT))) GLOMERULITIS)

(EXCUSE-FOR (RED-BLOOD-CELL-CASTS (PRESENCE ABSENT))
(HEMATURIA (SEVERITY (OR GROSS LIGHT))))

Diagram A-1: GLOMERULITIS SLICE



RELEVANT SYMPTOM — EVIDENCE/EXPECTATION VALUES

LGN:

HEMATURIA
  RECURRENT
  GROSS
  DURATION (LESS-THAN (YEARS 5.)) +STRONG
  DURATION (GREATER-THAN (YEARS 5.)) +WEAK

TIME-INDEX:
  (DURATION (BETWEEN (YEARS 1.) (YEARS 10.)) OFTEN)
  BIOPSY DIFFUSE-CHANGES -NECESSARY

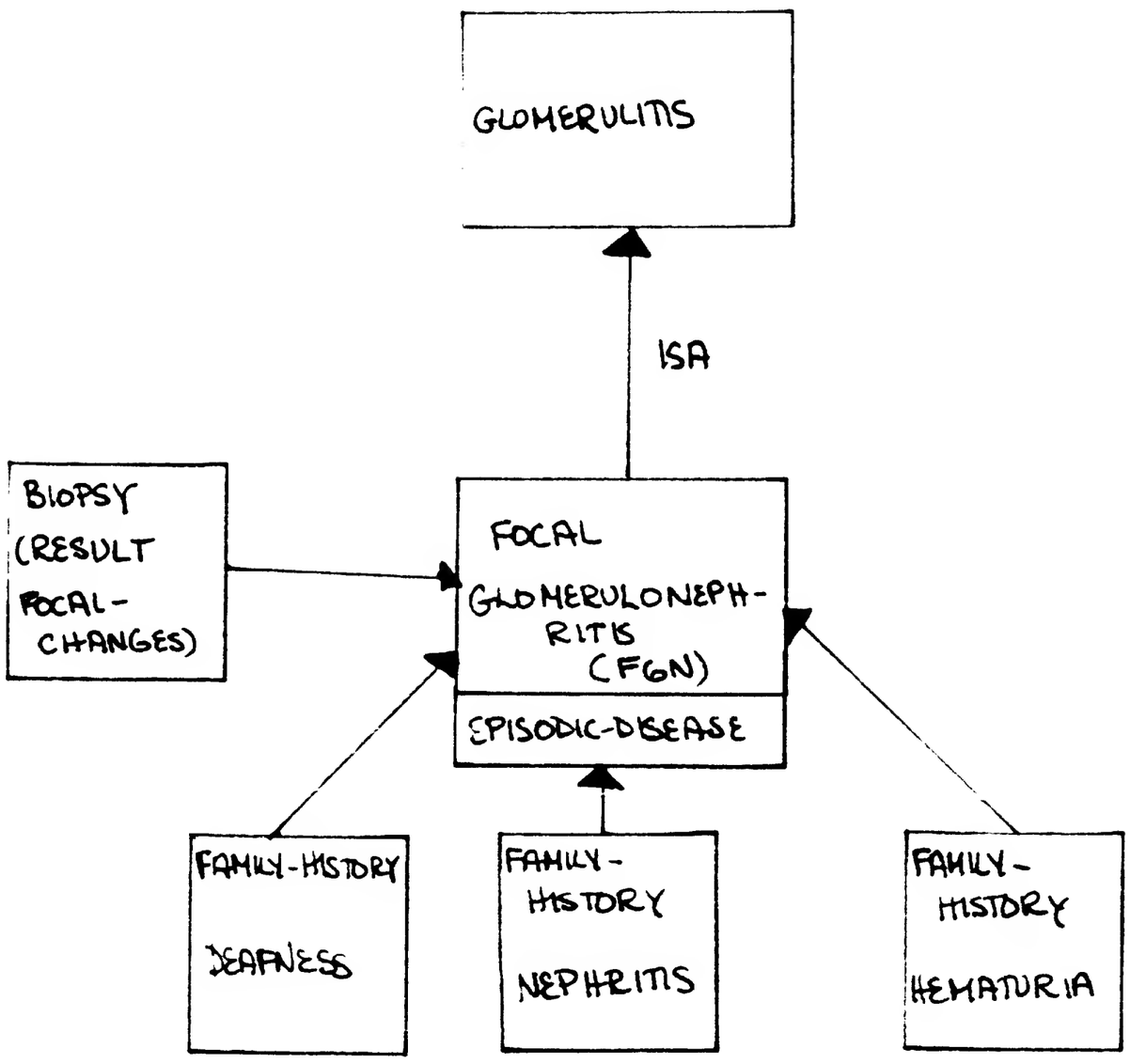
CGN:

HYPERTENSION
  CHRONIC +MODERATE

TIME-INDEX:
  (DURATION (BETWEEN (YEARS 0.) (YEARS 5.)) OFTEN)

(BEFORE AGN (START-TIME LGN.))
(BEFORE NEPHROTIC-SYNDROME LGN)

Diagram A-2: LGN AND CGN SLICES



RELEVANT SYMPTOM	EVIDENCE/EXPECTATION VALUES
FAMILY-HISTORY NEPHRITIS	† STRONG
BIOPSY FOCAL-CHANGES	-NECESSARY

(BEFORE (EPISODE FGN) UPPER-RESPIRATORY-INFECTION (INTERVAL (DAYS 1. 3.)))

TIME-INDEX:

(DURATION (BETWEEN (YEARS 0.) (YEARS 15.)) OFTEN)

Diagram A-3: FGN SLICE

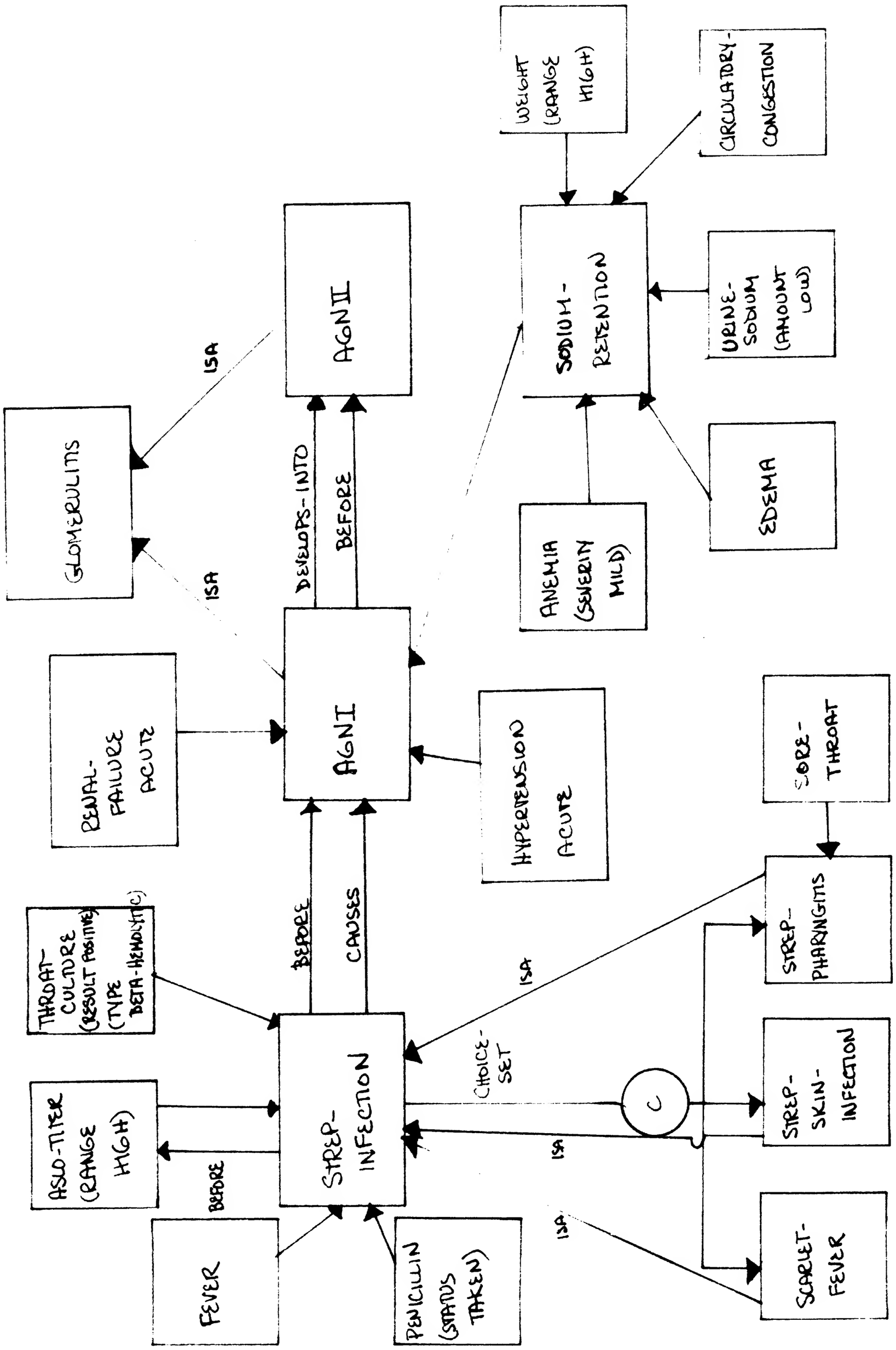


Diagram A-4 : AGN SUCC

STREP-INFECTION

FOR RELEVANT SYMPTOMS, SEE DIAGRAM 4-5.

(SUFFICIENT-CHOOSER SORE-THROAT STREP-PHARYNGITIS in
STREP-INFECTION)

(BEFORE STREP-INFECTION (ASLO-TITER (RANGE HIGH))
(INTERVAL (WEEKS 1. 5.)))

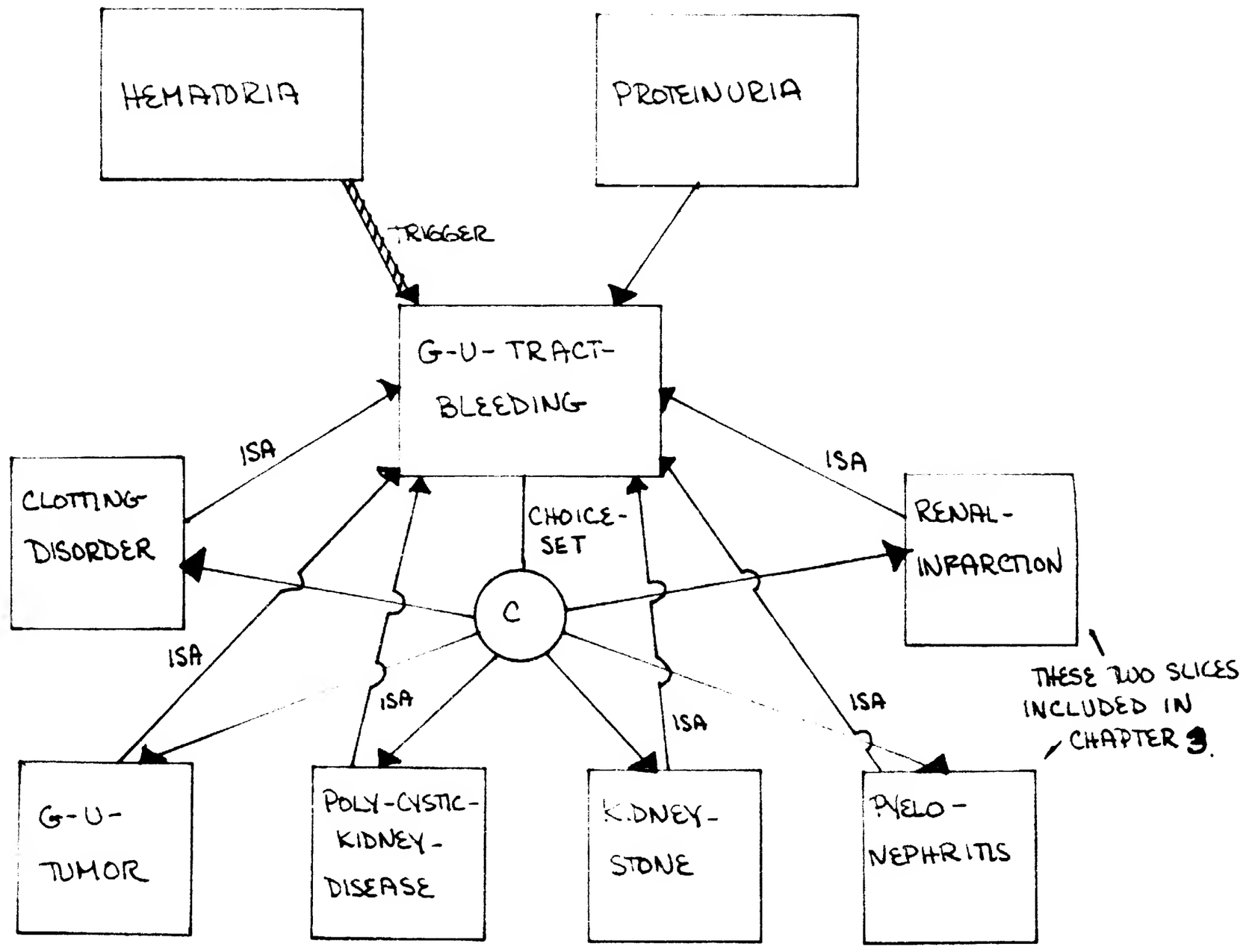
TIME-INDEX:
(DURATION STREP-INFECTION (WEEKS 1.) OFTEN)
(EXCUSE-FOR (ASLO-TITER (RANGE NORMAL)) (PENICILLIN (STATUS TAKEN)))

AGNI

TIME-INDEX:
(RECURRENCE NEVER)
(DURATION (INTERVAL (DAYS 3. 7.)) OFTEN)
(DURATION (GREATER-THAN (DAYS 7.)) SOMETIMES)

A-PRIORI-PROBABILITY:
((AGE CHILD) (SEX MALE) OFTEN)
(BEFORE STREP-INFECTION AGNI (INTERVAL (WEEKS 2. 3.)))

Diagram A-4a: AGN SLICE, CONTINUED.



(THIS CHOICE-SET IS INCOMPLETE)

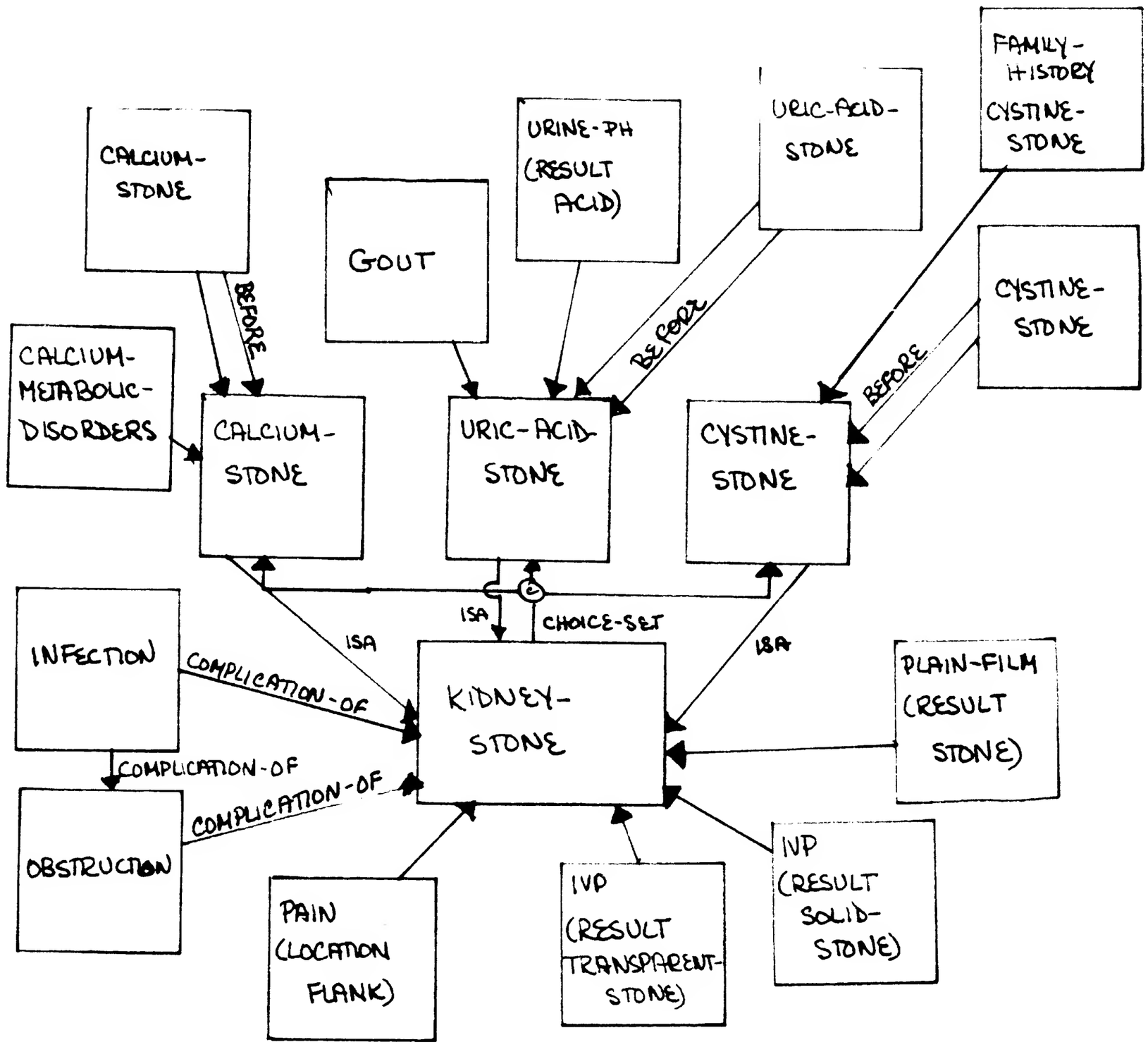
RELEVANT SYMPTOM	EVIDENCE/EXPECTATION VALUES
HEMATURIA	
GROSS	+STRONG
MICROSCOPIC	+MODERATE
ABSENT	-MODERATE
PROTEINURIA	
LIGHT	+STRONG

(PRECLUDES (AND (HEMATURIA LIGHT)
(PROTEINURIA HEAVY)) G-U-TRACT-BLEEDING)

G-U-TUMOR:

A-PRIORI-PROBABILITY
((AGE (OR YOUNG YOUNG-ADULT)) (SEX FEMALE) VERY-RARE)

Diagram A-5: G-U-TRACT-BLEEDING AND SOME EXAMPLES



RELEVANT-SYMPTOM EVIDENCE/EXPECTATION VALUES

IVP
TRANSPARENT-STONE +SUFFICIENT
SOLID-STONE +SUFFICIENT

PLAIN-FILM
STONE +SUFFICIENT

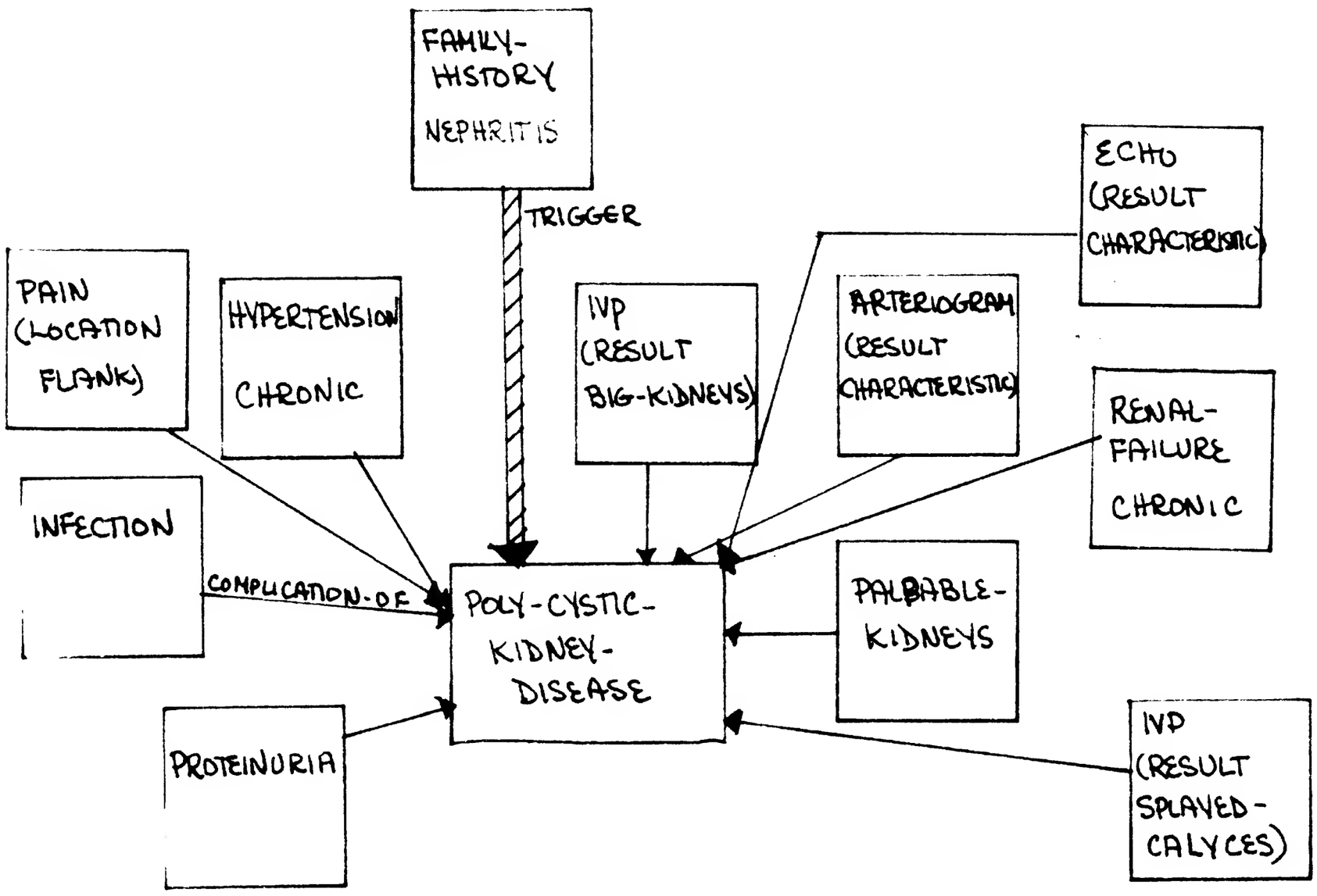
(SUFFICIENT-CHOOSER (IVP (RESULT TRANSPARENT-STONE)) URIC-ACID-STONE in KIDNEY-STONE)

(SUFFICIENT-CHOOSER (GOUT URIC-ACID-STONE in KIDNEY-STONE)

PAIN
(LOCATION FLANK) +STRONG
ABSENT -STRONG

TIME-INDEX:
(RECURRENCE (GREATER-THAN (YEARS 5.)) VERY-RARE)

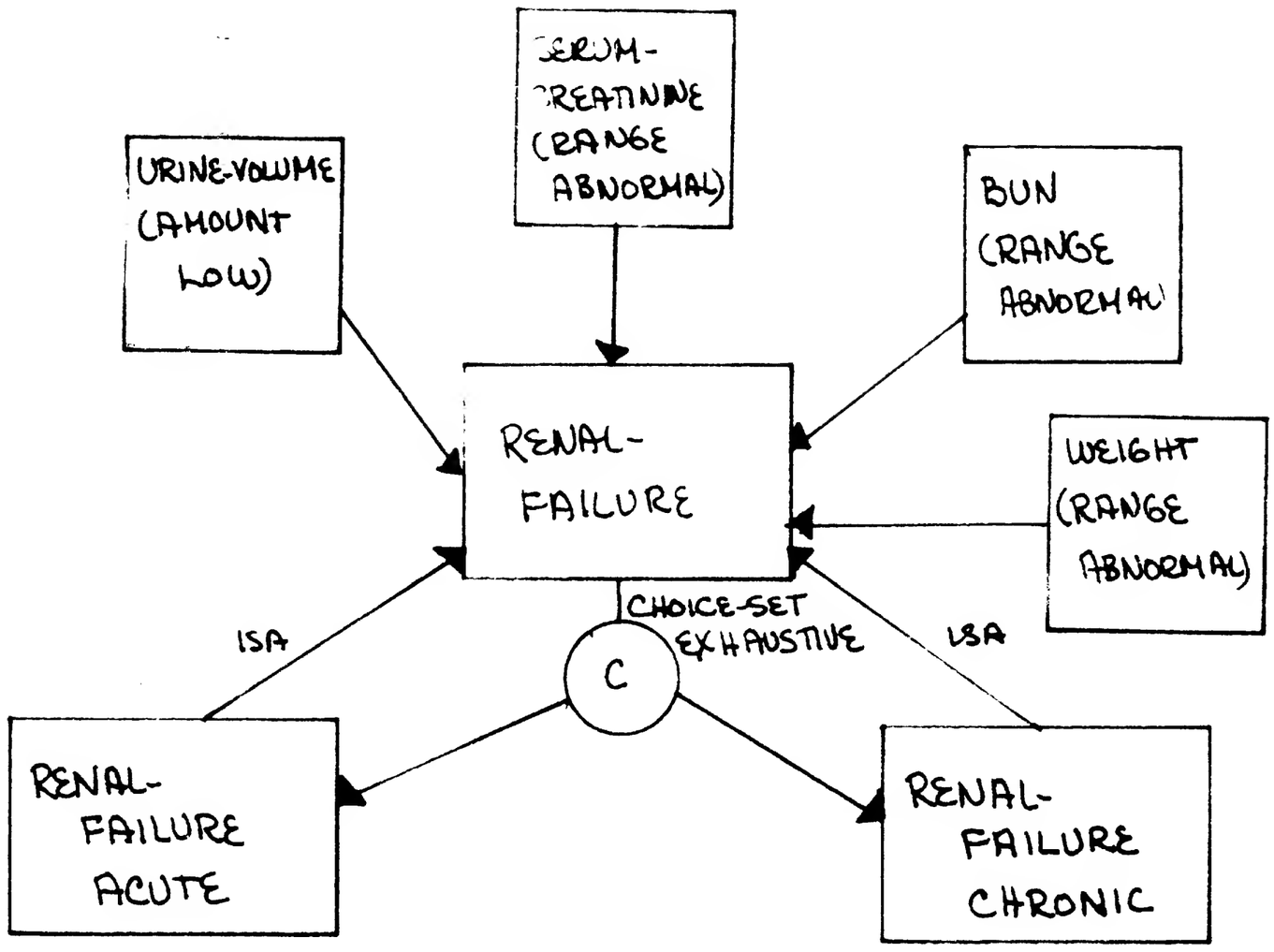
Diagram A-6: KIDNEY-STONE SLICE



RELEVANT SYMPTOM	EVIDENCE/EXPECTATION VALUES
FAMILY-HISTORY	
NEPHRITIS	+STRONG
PAIN	
FLANK	+WEAK
ABSENT	-WEAK
HYPERTENSION	
CHRONIC	+MODERATE
RENAL-FAILURE	
CHRONIC	+MODERATE
ABSENT	-STRONG
PALPABLE-KIDNEYS	
PRESENT	+STRONG
ABSENT	-STRONG

TIMEFINDEX: (DURATION (BETWEEN (YEARS 0.) (YEARS 15.)) OFTEN)

Diagram A-7: POLY-CYSTIC-KIDNEY-DISEASE SLICE



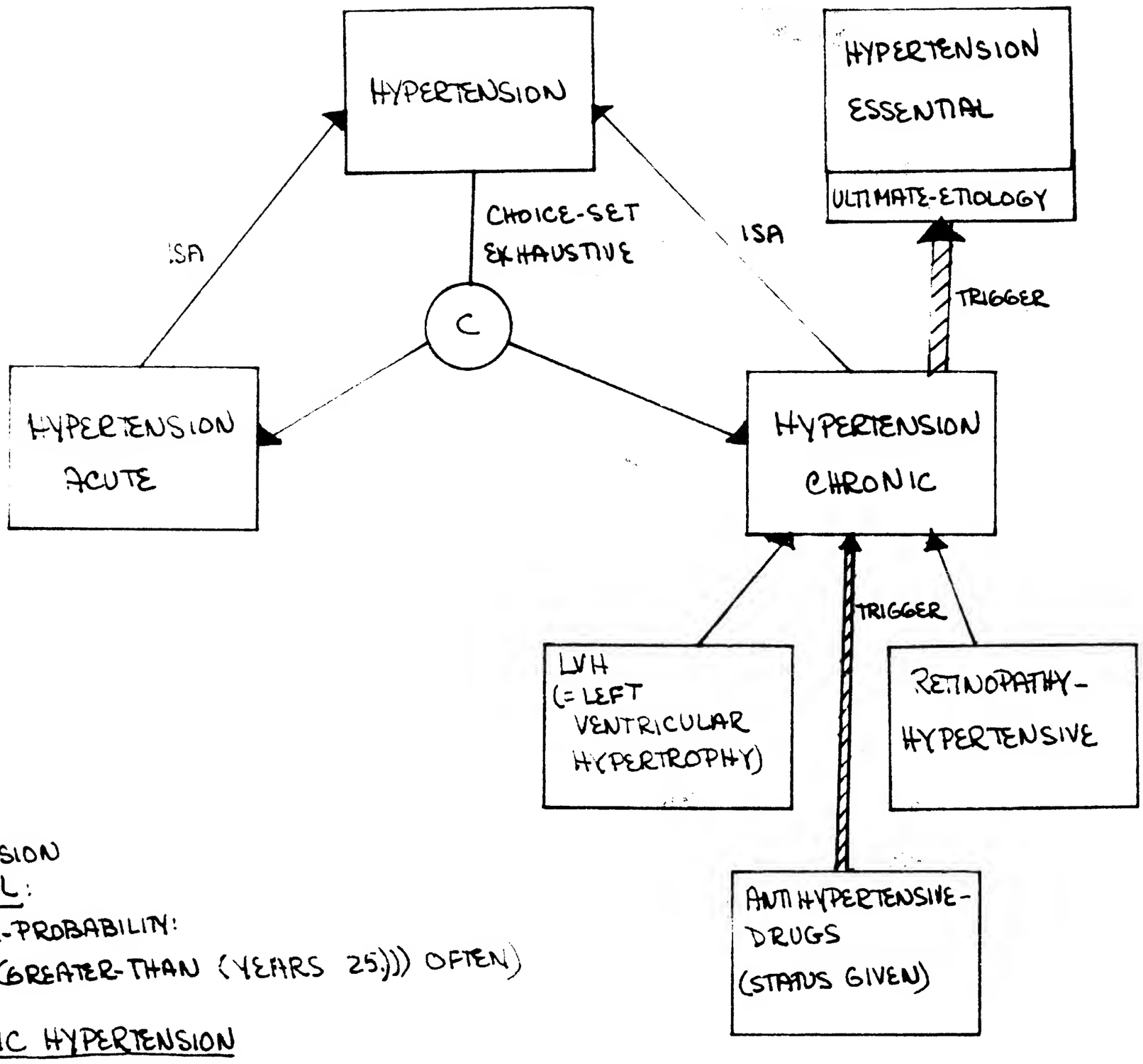
For relevant symptoms, see Diagram 4-5.

THE DISTINCTION BETWEEN ACUTE AND CHRONIC RENAL-FAILURE IS REALLY MADE ON THE BASIS OF TIME INFORMATION. NOT INCLUDED HERE.

```
(PRECLUDES (AND (BUN (RANGE HIGH))
                (SERUM-CREATININE (RANGE NORMAL))) RENAL-FAILURE)

(TRIGGERS (AND (BUN (RANGE HIGH))
               (SERUM-CREATININE (RANGE NORMAL))) NECROTIZING-TUMOR)
```

Diagram A-8: ACUTE AND CHRONIC RENAL-FAILURE SLICES



HYPERTENSION
ESSENTIAL:

A-PRIORI-PROBABILITY:
((AGE (GREATER-THAN (YEARS 25))) OFTEN)

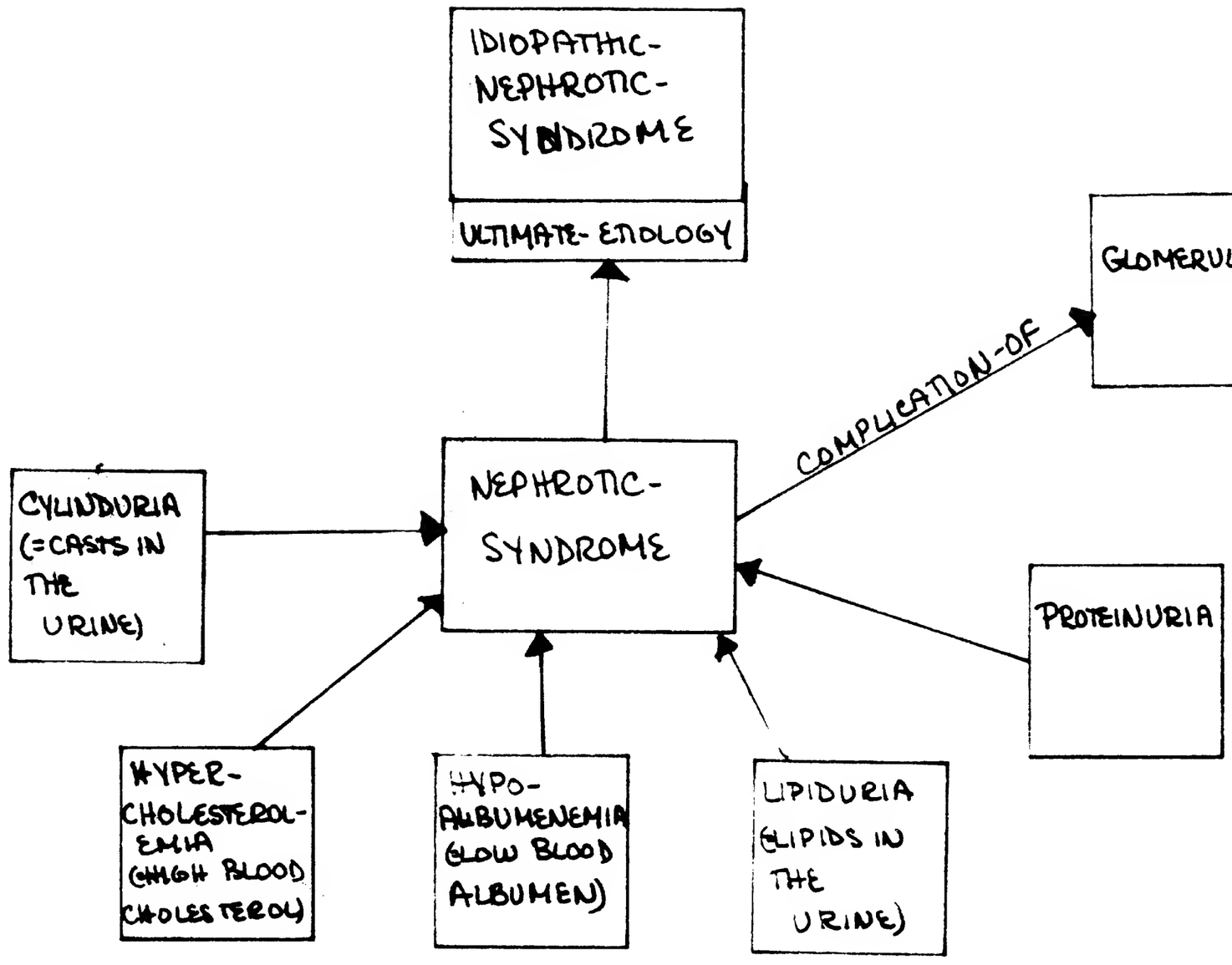
CHRONIC HYPERTENSION

RELEVANT SYMPTOM	EVIDENCE/EXPECTATION VALUES
LVH	+MODERATE
RETINOPATHY-HYPERTENSIVE	+MODERATE
ANTI-HYPERTENSIVE-DRUGS (STATUS GIVEN) (DURATION (GREATER-THAN (YEARS 1:)))	+SUFFICIENT

(SUFFICIENT-CHOOSER LVH (HYPERTENSION CHRONIC) in HYPERTENSION)
(SUFFICIENT-CHOOSER RETINOPATHY-HYPERTENSIVE (HYPERTENSION CHRONIC) in HYPERTENSION)

(EXCUSE-FOR (HYPERTENSION ABSENT) (ANTI-HYPERTENSIVE-DRUGS (STATUS GIVEN)))
(EXCUSE-FOR (HYPERTENSION ABSENT) (MYOCARDIAL-INFARCTION))

Diagram A-9: HYPERTENSION SLICE



RELEVANT SYMPTOM	EVIDENCE/EXPECTATION VALUES
PROTEINURIA	
HEAVY	+STRONG
LIGHT	-NECESSARY

Diagram A-10: NEPHROTIC-SYNDROME SLICE

CS-TR Scanning Project
Document Control Form

Date : 4/30/96

Report # AI-TR-316

Each of the following should be identified by a checkmark:
Originating Department:

☒ Artificial Intellegence Laboratory (AI)
☐ Laboratory for Computer Science (LCS)

Document Type:

☒ **Technical Report (TR)** ☐ **Technical Memo (TM)**
☐ Other:_____

Document Information

Number of pages: 244 (251-IMAGES)
Not to include DOD forms, printer intstructions, etc... original pages only.

Originals are:
☐ Single-sided or
☒ Double-sided

Intended to be printed as :
☐ Single-sided or
☒ Double-sided

Print type:
☐ **Typewriter** ☐ **Offset Press** ☐ **Laser Print**
☐ **InkJet Printer** ☐ **Unknown** ☒ **Other:**_____

Check each if included with document:

☒ **DOD Form** ☒ **Funding Agent Form** ☒ **Cover Page**
☐ **Spine** ☐ **Printers Notes** ☐ **Photo negatives**
☐ Other: _____

Page Data:

Blank Pages(by page number): _____

Photographs/Tonal Material (by page number): _____

Other (note description/page number):

Description :	Page Number:
IMAGE MAP!	(1-244)
UNNUMBERED TITLE PAGE, UNNUMBERED ABSTRACT, UNNUMBERED ACKNOWLEDGEMENTS	4-244
(245-251) SCANCONTROL, COVER, FUNDING AGENT, DOD, TRGT'S (3)	

Scanning Agent Signoff:

Date Received: 4/30/96 Date Scanned: 5/6/96 Date Returned: 5/9/96

Scanning Agent Signature: Michael N. Cook

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AI-TR-316
2. GOVT ACCESSION NO.:
3. RECIPIENT'S CATALOG NUMBER:

4. TITLE *(and Subtitle)*: Hypothesis Formation and Evaluation in Medical Diagnosis

5. TYPE OF REPORT & PERIOD COVERED: Technical Report

6. PERFORMING ORG. REPORT NUMBER:

7. AUTHOR(s): Ann D. Rubin

8. CONTRACT OR GRANT NUMBER(s): Nooo14-70-A-0362-0003

9. PERFORMING ORGANIZATION NAME AND ADDRESS:
Artificial Intelligence Laboratory
545 Technology Square
Cambridge, Massachusetts 02139

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:

11. CONTROLLING OFFICE NAME AND ADDRESS:
Advanced Research Projects Agency
1400 Wilson Blvd
Arlington, Virginia 22209

12. REPORT DATE: January 1975

13. NUMBER OF PAGES: 246

14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)*:
Office of Naval Research
Information Systems
Arlington, Virginia 22217

15. SECURITY CLASS. *(of this report)*: UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT *(of this Report)*:

Distribution of this document is unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*:

18. SUPPLEMENTARY NOTES:

None

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*:

Artificial Intelligence
Computer-aided medical Diagnosis
Expert problem-solving
Heuristic methods
~~Hypothesis~~ evaluation
Hypothesis formation
Local vs. global processes
Protocol analysis

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*:

This thesis describes some ~~aspects of a computer system~~ for doing medical diagnosis in the specialized field of kidney disease. Because such a system faces the spectre of combinatorial explosion, this discussion concentrates on heuristics which control the number of concurrent hypotheses and efficient "compiled" representations of medical knowledge.

(continued on reverse side)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

Scanning Agent Identification Target

Scanning of this document was supported in part by the **Corporation for National Research Initiatives,** using funds from the **Advanced Research Projects Agency** of the **United states Government** under Grant: **MDA972-92-J1029.**

The scanning agent for this project was the **Document Services** department of the **M.I.T Libraries.** Technical support for this project was also provided by the **M.I.T. Laboratory for Computer Sciences.**